# The Role of Motivation in Credibility Evaluations of Online Information

Faculty of Behavioral Sciences
Masterthesis Psychology
University of Twente

Ewald T.A. Maas
19-06-2013

Examination Committee
Dr. Matthijs L. Noordzij
Prof. Dr. Jan Maarten C. Schraagen

This report is dedicated to my late grandmother.

**Abstract**

This study set out to explore the role of motivation in credibility evaluations of online information. The experiment conducted to achieve this goal, included successfully manipulating the amount of motivation of participants after which they performed an online information searching task. Activities and thoughts of participants were elicited through screen capture recordings that served as visual cues for a retrospective think-aloud exercise. After performing a template analysis of the verbatim it was found that 'barely motivated' participants did not perform any meaningful credibility evaluations. Conversely, 'strongly motivated' participants showed considerable credibility concerns and performed several credibility evaluations. These credibility evaluations were all of a heuristic nature; no systematic credibility evaluations occurred. Instead, it seems that 'strongly motivated' participants used another heuristic to making systematic credibility evaluations obsolete: 'searching only in scientific search engines'. These results question the extent to which Metzger's proposed dual-processing model of credibility assessment holds ground.

*Keywords*: *motivation, credibility evaluation, online, dual-processing*

Deze studie onderzocht in welke mate motivatie de vorming van online betrouwbaarheidsoordelen stuurt. Het experiment om dit te bepalen manipuleerde de motivatie van de deelnemers waarna ze een online informatie zoek taak uitvoerden. De activiteiten en gedachten van de deelnemers werden vastgelegd door middel van beeldscherm opnames tijdens de informatie zoek taak , die tevens dienden als visuele ondersteuning voor een retrospectieve hard-op denken oefening. Door middel van een 'template' analyse van de verbalizaties bleek dat amper gemotiveerde proefpersonen amper noemenswaardige betrouwbaarheidsevaluaties uitvoerden. Tevens bleek dat zeer gemotiveerde proefpersonen vaak bezorgd waren over de betrouwbaarheid van de informatie en daartoe betrouwbaarheidsevaluaties uitvoerden. Hiertoe gebruikten ze verscheidene heuristiek en viel verder op dat dit niet op een systematische manier gebeurde. Sterker nog, het lijkt alsof de zeer gemotiveerde deelnemers een heuristiek gebruikten om zeker te weten dat informatie betrouwbaar was, te weten: 'online zoeken beperken tot wetenschappelijke zoekmachines'. Deze resultaten zetten vraagtekens bij het model van Metzger dat systematische betrouwbaarheidsevaluaties voorspelt voor gemotiveerde informatie zoekers.

*Trefwoorden: motivatie, evaluatie van betrouwbaarheid, online, dual-processing*

## Contents

# The Role of Motivation in Credibility Evaluations of Online Information

## Introduction

Try to remember the last time that you visited a website and read information on that website. To what degree did you consciously evaluate the reliability of the information you read? Did you ever navigate away from a website just because it did not 'look right'? Do you frequently navigate to the same website, because it has provided you with reliable information in the past? What makes you (dis)trust the information that you read online? These questions are related to the research domain of online credibility evaluations. *One specific question that this study explores is how motivation influences online credibility evaluations during the information seeking process.*

## Background

During the 20th century a series of technological advances have caused a change in our appetite for information. Traditional forms of media such as books, newspapers, radio and television gained a new platform to reach the consumer: the internet. The transformation (and perhaps transition) of traditional media to their digital counterparts is becoming more profound. Books are now available through the internet on e-readers, newspapers are read on iPads, the list of internet radio stations is limitless and television channels can be customized using Google TV. An essential factor that enables these transformations is accessibility to the internet. In the late 20th century, internet by use of landlines began permeating society. Seybert (2011) reports that in 2011 internet penetration in households reached over 90% in six European countries and that other European countries were not far behind. This means that virtually anyone who wishes to 'go on the internet' has the means to do so. With increasing connection speeds on wireless devices and greater wireless coverage, this also means that *anyone* can access the internet *anywhere* at *any time*.

Connecting someone to the internet also means introducing a potential producer of information to the network. Traditionally, communicating to larger crowds through media was only available for professional institutions e.g. television networks, newspapers. Today, barriers that kept people from communicating with each other dissolve more and more with innovations as blogs, wiki's and social media. Shirky (2008) provides a detailed account of the

transformations of the media landscape during the last century and states "Our social tools remove older obstacles to public expression, and thus remove the bottlenecks that characterized mass media. The result is the mass amateurization of efforts previously reserved for media professionals" (p.54). This increases the pressure on the information seeker to evaluate the credibility of the information that he[1] finds on the internet. An illustrative example is the collaborative encyclopedia named Wikipedia that is put together by anyone who wishes to contribute his knowledge. A page on Wikipedia can be as accurate and extensive as the traditional Encyclopedia Britannica, but due to its very nature of open-source collaboration the opposite can be true for any given page as well. *How is the information seeker going to tell the difference in quality of these pages and how does the information seeker evaluate the credibility of the information he reads online?*

**Credibility**

The concept of credibility dates back to Aristotle who considered a speaker's ethos to be the credibility or trustworthiness of the speaker. This aspect of credibility is now referred to as *source credibility* and has been extended to cover written material too. In a study performed by Kubiszewski, Noordewier, and Costanza (2011) the effects of source credibility are clearly visible. Participants in their research read an article on climate change on Wikipedia, Encyclopedia of Earth or Encyclopedia Britannica and rated articles from Encyclopedia Britannica more credible than articles from Wikipedia or Encyclopedia of Earth even though the content of the articles was exactly the same.

Besides source credibility it is also the content itself that determines the credibility perception of the information. For instance, Dutta-Bergman (2004) showed that incompleteness of health information online was found to negatively impact the credibility evaluation. In addition, a study by Lucassen and Schraagen (2011) suggests that content with varying accuracy influenced the credibility evaluations of domain experts.

Even though the information content and source can seem credible, the manner in which the information is presented (e.g. writing style, references) also influences the credibility evaluation (Lucassen & Schraagen, 2011). In a study performed by Fogg et al. (2003) participants performed credibility assessments of two different websites. The design of the website was

[1] For improved readability he/she and his/her will be shortened to 'he' and 'his' respectively for the remainder of this thesis.

mentioned in 46,1% of the comments, indicating that it is a prominent constituent of the overall evaluation.

It becomes clear that Aristotle's idea of credibility has been expanded beyond source credibility. Not only does the source of the information influence the credibility evaluation of the information seeker, but the information content and the way the information is presented also seem to have a profound effect.

**Integrated model of trust in online information**

The degree to which the characteristics of the information (source, content, presentation) influence the credibility evaluation is mediated by the user characteristics of the information seeker (Lucassen & Schraagen, 2011). One information seeker may have more domain expertise on the subject matter and will therefore have a better chance of detecting factual errors. Another might have read information from the same source before and will therefore be biased in evaluating the credibility of the source. In a similar fashion, a certain layout or writing style might correspond to the personal preferences of one information seeker but not the other. In his doctoral dissertation Lucassen (2013) has addressed the influence of user characteristics on the formation of credibility evaluations and proposed the *integrated model of trust in online information* as shown in Figure 1. Among other things this model states that a domain expert will be able to evaluate the semantic features of the information e.g. factual accuracy. The result of this credibility evaluation will then influence the trust in the information.

The fact that the user characteristics determine the credibility evaluation behavior is in accordance with previous research and theories. The prominence-interpretation theory as proposed by Fogg (2003) states that an item of information first needs to be prominent for the information seeker in order to be evaluated. Whether an item of information is prominent to an information seeker depends on his characteristics. An information seeker with high information skills will for instance recognize the reference list as an indicator of credibility after which he will interpret the references that will then influence the credibility evaluation. The integrated model of trust in online information also fits in similar fashion with the unifying framework of credibility assessment as proposed by Hilligoss and Rieh (2008). In their model the personal definition (i.e. the information seeker's definition) of trust determines how one evaluates the credibility of information. An example of such a personal definition of trust is an information seeker that has been taught that an objective writing style indicates high quality information and

is therefore trustworthy. Whenever this information seeker examines information he will value the writing style and use it as an indicator for the credibility of the information.

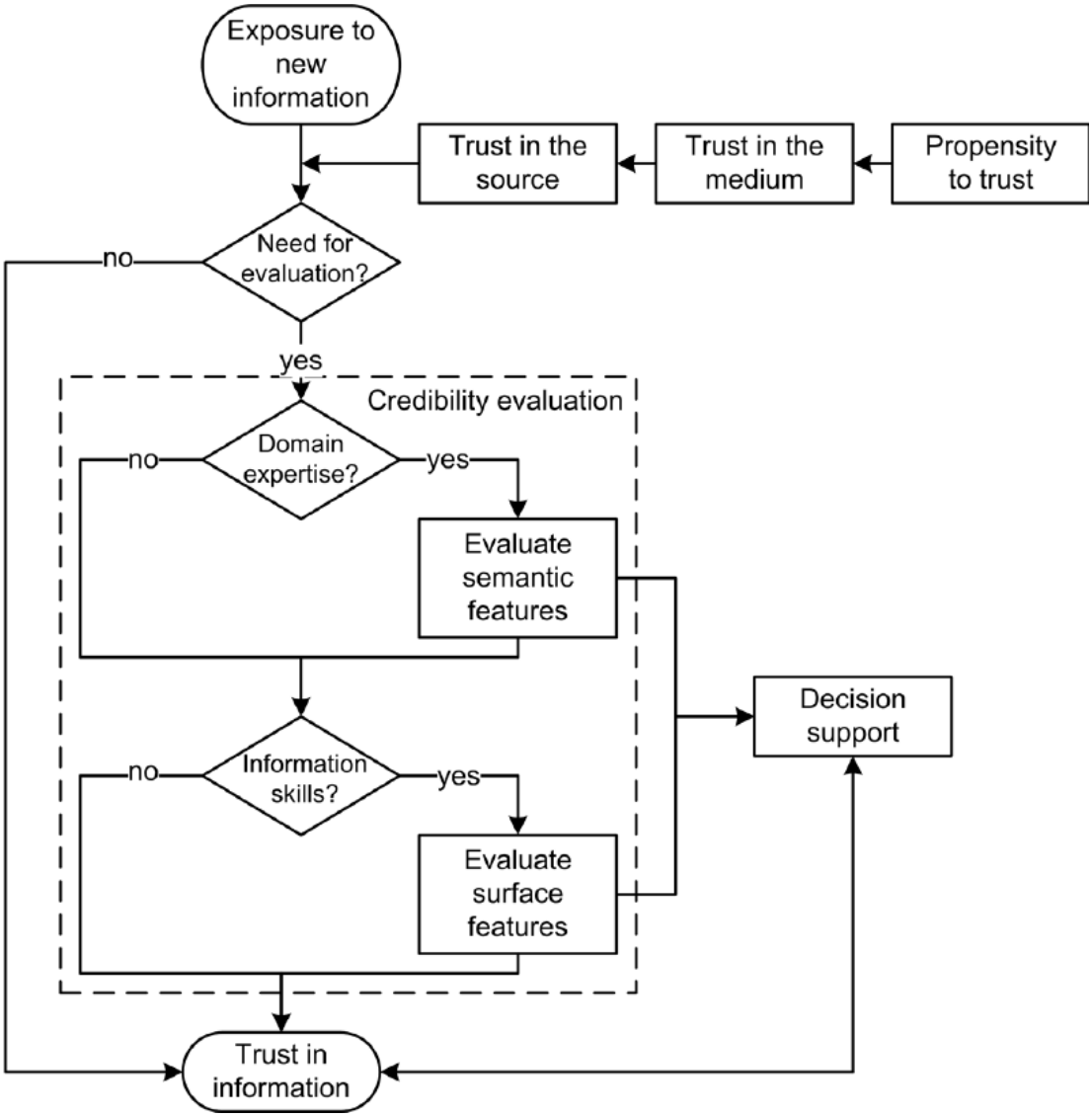


*Figure 1.* Integrated model of trust in online information as presented in Lucassen (2013).

**Dual-Processing approach to credibility evaluations**

The integrated model of trust in information relates which user characteristics enable the information seeker to evaluate the semantic and surface features of information. The manner in which the able information seeker performs the evaluation of the semantic and surface information features is often cited to follow a dual-processing approach as proposed by Metzger (2007).

***Heuristic-Systematic Model.*** Metzger proposed that the dual-processing model of information processing developed by Chaiken (1980) and labeled the *Heuristic-Systematic Model* (HSM) can be used to characterize the online credibility evaluation process. Chen and Chaiken (1999) provide a comprehensive view of the workings of the HSM:

> Within any given judgmental context, the heuristic-systematic model delineates two basic modes by which perceivers may determine their attitudes and other social judgments. *Systematic processing* entails a relatively analytic and comprehensive treatment of judgment-relevant information. (...) Given its nature, systematic processing requires both cognitive ability and capacity. (...) The other basic mode, *heuristic processing*, entails the activation and application of judgmental rules or "heuristics" that, like other knowledge structures, are presumed to be learned and stored in memory (e.g. "Experts' statements can be trusted," "Length implies strength," "Consensus opinions are correct"). (...) Relative to systematic processing, heuristic processing makes minimal cognitive demands on the perceiver. (p. 74)

Chen and Chaiken further explain that perceivers follow 'the sufficiency principle' meaning that cognitive effort is exerted until a person's actual confidence level in their judgment reaches their desired level of confidence or sufficiency threshold. Whether heuristic or systematic processing takes place is guided by a 'principle of least effort'. Heuristic processing will therefore predominate over the generally more effortful systematic processing. When the gap between the actual and desired level of confidence cannot be met by heuristic processing alone, the perceiver might engage in systematic processing to close the gap.

***HSM applied to online credibility evaluations.*** Metzger (2007) proposed to use the HSM as a basis for the online credibility evaluation process as seen in Figure 2. An example of how the HSM might be applicable to credibility evaluations can be found in the study by Hilligoss and Rieh (2008) when a students' information searching strategy is mentioned. The student goes through a series of credibility evaluations until she has enough confidence in the information she received. This process closely represents the aforementioned sufficiency principle of exerting cognitive effort until the actual level of confidence meets the desired level of confidence. When applying the HSM in online credibility evaluations, the difference for example between evaluating surface features of information in a heuristic or systematic manner is the difference

between glancing at the amount of references (heuristic) as in indicator for credibility versus inspecting each reference to be valid (systematic).

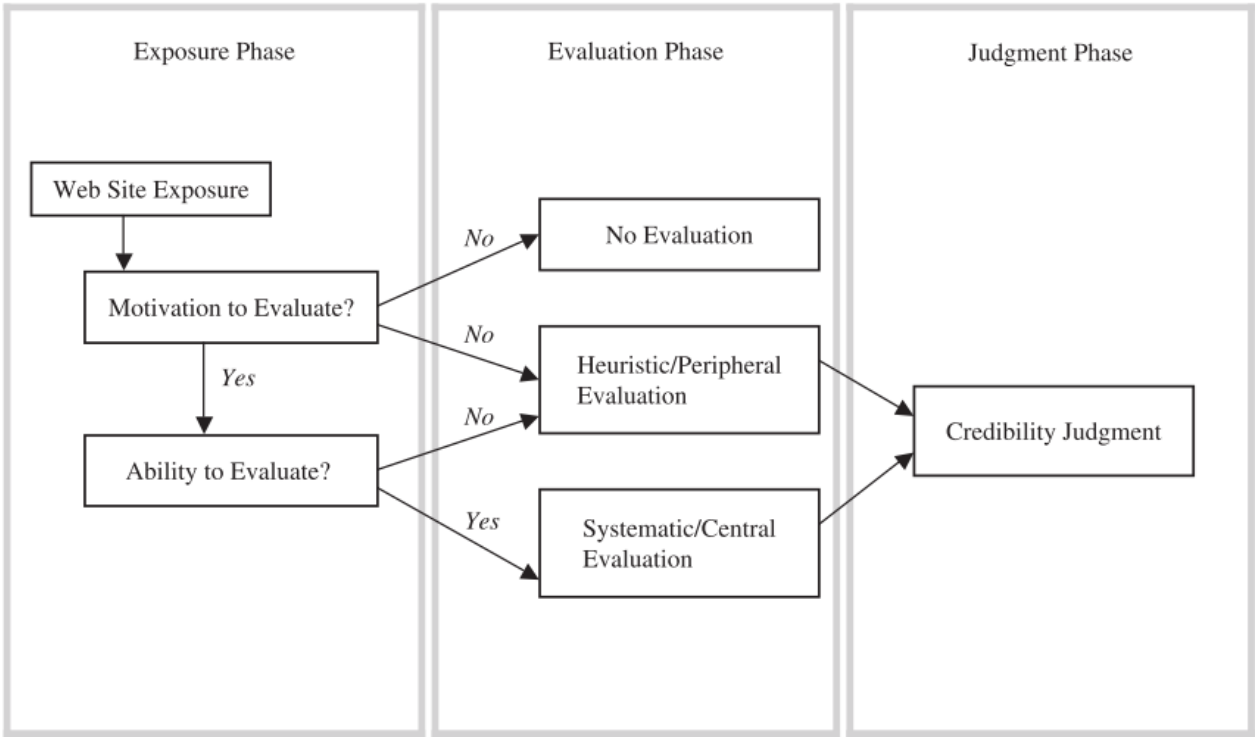


*Figure 2.* Dual-Processing model of credibility assessment as depicted in Metzger (2007).

In Metzger's model in Figure 2 it are motivation and ability that codetermine which type of evaluation is likely to happen. Metzger (2007) states that "Motivation comes from the consequentiality of receiving low-quality, unreliable, or inaccurate information online" (p. 2087). This, according to the HSM, would thus mean that low-motivation does little to raise the desired confidence level. In a similar analysis, high motivation would require a higher level of desired confidence, since the consequentiality of receiving low-quality, unreliable or inaccurate information would be higher. Following this train of thought connecting the HSM to motivation in credibility evaluations seems to suggest that with higher motivation come higher desired levels of confidence which will elicit more systematic processing. This as opposed to lower levels of motivation, with lower desired levels of confidence which can be met with heuristic processing.

**Support for Metzger's Dual-Processing approach?**

The proposed model of Metzger (2007) clearly suggests the role that motivation plays in credibility evaluations of online information. Together with the information seekers' ability it

explains what type of processes (i.e. heuristic or systematic) can be expected to occur when evaluating semantic and surface features of online information. Despite research (Stanford, Tauber, Fogg, & Marable, 2002; Lucassen & Schraagen, 2011; Lucassen, Muilwijk, Noordzij, & Schraagen, 2013) on the influence of the information seeker's ability on credibility evaluations, there is little research to support the suggested role of motivation. Metzger (2007) mentions several studies that–according to her–provide preliminary evidence for the suggested role of motivation.

First, Metzger (2007) interprets the results of a previous study by Flanagin and Metzger (2000) as support for motivation's role in online credibility evaluations and states:

> (...) credibility assessment varied by information type, such that participants reported exerting significantly more effort to verify the credibility of reference information (defined as factual, non-news information such as self-help, health-related, or "how-to" information) than they did for entertainment information. (p. 2089)

However, the word 'motivation' is not mentioned in the study of Flanigan and Metzger (2000). The study asked the participants about their verification behavior with regard to information depending on the information type. Four information types were distinguished: factual, commercial, news, entertainment. In laymen's terms the results indicate that people verify the information received differently depending on the information type. It is pure speculation to conclude that these findings support the role of motivation in credibility assessment. This was not tested nor accounted for and neither was it done in the follow-up study performed by Flanagin and Metzger (2007), which Metzger (2007) also cites as a reference supposedly supporting the role of motivation in online credibility evaluations.

Second, Metzger (2007) also relates the study of Stanford et al. (2002) to support her idea that motivation is related to credibility assessment. It is true that in this study the ability of the reader (expert vs. consumer) resulted in different credibility judgments. However motivation or strongly related concepts such as task importance or purpose were not mentioned in this study.

On closer inspection it seems speculative to interpret the findings of these studies as preliminary evidence for the suggested role of motivation in online credibility evaluations. There are some other studies, however, that seem more promising in this regard even though these studies show tentative support at best.

In a study performed by Dutta-Bergman (2004) participants visited a website with information regarding heart disease. One group of participants was either told to browse the website for information at their own lenience without any further instruction. Another group of participants had to imagine that they suffered from heart disease and that they should search the website for information regarding their condition. The results of this study showed that participants who were browsing without any goal directedness were less likely to judge the source of information (i.e. the website) as more credible compared to the participants who were searching for information regarding their own make belief heart disease. Unfortunately, the research scope of the study by Dutta-Bergman (2004) was mainly focused on information completeness rather than motivation and focused solely on source credibility. It was not able to provide an analysis why the credibility evaluations were lower or higher in the respective cases.

A study by Rieh and Belkin (1998) did relate motivation to credibility concerns. They found that participants were more concerned with the credibility of information when the information was expected to have a greater impact on their lives. Similar concerns for credibility were expressed in the study by Hilligoss and Rieh (2008). In their study participants were more concerned with the credibility of information when it involved personal information needs such as health and finances, academic achievement, and problem solving, as opposed to searching information for entertainment purposes. Participants were also more concerned with credibility when searching for information for someone else. Both studies clearly relate goal-importance, which can be seen as a form of motivation, to credibility concerns, but do not provide a thorough analysis. It is for instance unclear if more credibility concerns lead to different credibility evaluation processes (i.e. heuristic or systematic), which the dual processing model of Metzger (2007) predicts.

The study of Chaiken and Maheswaran (1994) does relate task importance in an off-line setting to specific kinds of processing. They demonstrated that high task importance can lead to systematic processing and that low task importance always leads to heuristic processing. This study supports the HSM as proposed by Chaiken (1980), yet it remains unclear if the findings can be applied to online credibility evaluations when an information seeking task is concerned.

**Research scope**

With regard to the reviewed literature it can be stated that the application of the HSM to online credibility evaluations as proposed by Metzger is primarily theoretical. The few empirical

studies that are related to motivation and its suggested role in online credibility evaluations show some tentative support, but fail to provide a comprehensive understanding. *The research goal of this study is therefore to explore the role that motivation plays with regard to credibility evaluations of online information.*

The 'integrated model of trust in online information' by Lucassen (2013) and the 'dual-processing model of credibility assessment' by Metzger (2007) use exposure to new information as the first step in their models. An information seeker's motivation, however, is present as soon as the information seeking task starts i.e. before the actual presence of new information. It is therefore easy to imagine that 'a desired level of confidence' from the HSM is already established before the actual searching for online information is started. Furthermore, if the information seeker is determined to find only very reliable information, it is also easy to imagine that it affects the search engine selection process such as selecting Google Scholar (i.e. where the search results in academic output) over Google Search. This choice–partly because of motivation–would then impact all future search results and consequentially all future credibility evaluations. The possible early impact of motivation with regard to future credibility evaluations can therefore not be ignored. This study will therefore not set out to 'test' the validity of Metzger's dual-processing model for credibility assessment within the boundaries of a single webpage. Instead it will explore to what degree motivation impacts the formation of online credibility evaluations *within the entire information seeking process.*

## Experiment design considerations

This section will first address how studying motivation within the context of the entire information seeking process impacts the experiment design setup. Then will be explained why participant selection is closely tied to the search task selection and information needs. After selecting participants, a pilot study will be conducted in which the search tasks will be selected that are to be used in the main experiment of this research. This section concludes with clarifying choices regarding the procedure and data analysis of the main experiment.

### Experiment design

The ideal experiment for this study would observe its participants while they are performing a real-life information searching task and in some way unobtrusively elicit their motivation, credibility evaluation processes and credibility evaluations. Unfortunately it would be

difficult, if not impossible within the time limitations of a master thesis, to relate differences in credibility evaluations and credibility evaluation processes to motivation, because the goal of each search task and the amount of motivation related to it would differ in each setting of each participant. Conclusions are then expected to be limited and similar to Hilligoss and Rieh (2008) i.e. each different search task generates an amount of motivation with varying credibility concerns. Therefore, this study will attempt to manipulate the amount of motivation of the participants by giving them a *specific search task* that falls within *a specific information need*. The given search task should be a familiar search task for the participants, so that motivation corresponding with the search task is reflective of day-to-day motivation of the participants. Resulting differences in credibility evaluations and related processes are then expected to be due to the difference in motivation of the participants and can be readily compared.

**Information needs of online information seekers**

Several studies have provided a taxonomy of the information needs of online information seekers (Broder, 2002; Rose & Levinson, 2004; Kellar, Watters, & Shepherd, 2006). Kellar et al. (2006) for instance distinguish the following information needs based on the internet use of 21 university students: specific fact finding (18.3% of total web use), information gathering (13.5%), undirected browsing (19.9%), web-mediated activities such as checking email (46.7%) and other activities (1.7%). It can be expected that these percentages change over time as the world wide web evolves, and that the percentages are different for different sub-populations e.g. university students might spend more time gathering information than elderly people. The information need on which this study is focused is therefore not predetermined, but its selection follows from the selection of the participants and their sub-population specific information needs and related search tasks. This indicates that participant selection is closely tied to the possible information need and search task selection for this study.

**Participant selection**

The information search task to be performed by the participants is expected to motivate the participants to a certain degree. The motivation that will result from the information search task will partly be determined by the type of search task and its topic. The greatest difference in motivation is therefore to be expected when one group of participants is discouraged by the type of search task and topic while another group of participants is motivated by the type of search task and topic. This has two implications for the participant selection. First, in order to generate

authentic motivation (or lack thereof) it is required to select a search task that is as authentic as possible for all participants. This therefore requires knowledge of the participants' daily search tasks. Second, in order to generate motivation (or lack thereof) participants are required to have a shared interest or disinterest in a certain topic or field.

For these reasons 1st and 2nd year psychology students of the University of Twente (in Enschede, The Netherlands) were selected. Psychology students perform various information search tasks during their studies. It is for this reason that the information need 'information gathering' was selected, since typical information gathering tasks that students perform are known to the researcher. Also, psychology students are guaranteed to have at least one shared interested i.e. psychology. Based on the curriculum of the students it is known which topics with regard to psychology are familiar to the students and are likely to interest and motivate them. Furthermore, 1st and 2nd year psychology students are expected to have similar information searching skills and similar information skills, which can therefore be excluded as possible confounding variables in this study. The reason why they are expected to have similar skills in this regard is because of a mandatory course called 'academic competences' that educates them in these manners. In addition, 1st and 2nd year psychology students receive course credits when they participate in psychological research, which eases availability of participants for this study.

**Search task selection: pilot study**

In order to select search tasks and topics that motivate 1st and 2nd year psychology students a pilot study was conducted.

***Participants.*** Emails were sent and tracked using MailChimp (an email marketing provider) to 387 psychology students enrolled in 1st and 2nd year psychology courses at the University of Twente that asked them to complete an online survey. 116 students opened the email, which led to 66 surveys responses. 14 incomplete surveys were removed, in which none of the scenario questions were answered. From the remaining 52 surveys, 2 surveys were removed due to the relative old age (47 and 50 years) of the participants. 50 surveys were completed or partially completed of which the average age of the participants was 21.5 years (SD=2.2). 31 Dutch and 19 German psychology students participated (41 female, 8 male, and 1 undefined). Participants did not receive any compensation for taking part in the online survey.

***Search tasks.*** Students were either currently enrolled in or had already completed the course 'academic competences'. In this course students are taught for instance how to write essays,

the value of references etc. Based on the teaching goals and curriculum of this course, 8 scenarios were developed (see Appendix A-I) that closely resembled typical assignments from the 'academic competences' curriculum. To this end, students were also told to pretend the scenarios were part of the course 'academic competences'. Topics (also listed in Appendix A-I) that the scenarios featured were chosen to be either study related and were therefore expected to interest and motivate them, or to be completely unrelated to their studies and therefore expected to be discouraging.

***Design and variables.*** Independent variables were the search task as seen in the 8 scenarios and the topic of the search task, which were either supposedly interesting psychology related topics designated with (+) or supposedly uninteresting topics designated as (-). This resulted in a 8x2 between subjects design. Dependent variables measured on 7 point Likert scales are motivation, topic familiarity and interest in topic. The first point on the scale indicated 'not motivated', 'not familiar' and 'not interested' respectively. The seventh point on the scale indicated 'highly motivated', 'highly familiar' and 'highly interested'.

***Procedure.*** Participants were contacted through email which contained a hyperlink to the online survey. After filling out personal information, participants proceeded to the main body of the survey. Each participant was asked to answer questions related to 4 scenarios of which an example is seen in Appendix A-II. Scenarios that were presented to the participants constituted of 2 scenarios with topics that were expected to interest them (+) and 2 scenarios with topics that were not expected interest them (-). Every participant was randomly assigned to one of the four groups as seen in Appendix A-I, Table I, which indicate the scenario and topic order. Scenarios and topics were pseudo-randomly selected for each group so that participants were never shown the same scenario or topic twice. Furthermore the order of scenarios and their interesting (+) and uninteresting (-) topics in these groups were pseudo-randomly selected to counteract possible 'order' effects.

***Results.*** Participants were expected to complete 200 scenarios in total (50 participants x 4 scenarios) and through pseudo-randomization described above to be treated as independent samples. 11 scenarios were not completed due to unknown reasons. Presumably participants were no longer interested in completing the survey or completing the survey took longer than expected. Mean motivation scores are listed for each scenario and topic combination in Table 1. Motivation scores were highest for scenarios C(+) (M=5.2, SD=1.2) and E(+) (M=5.1, SD=1.4) and was

lowest for scenario E(-) (M= 2.0, SD=1.2). The difference in motivation score between E(+) and E(-) is significant ($F$ (1, 20) = 28.84, $p < 0.01$), this is also the case for C(+) and E(-) ($F$ (1, 16) = 31.74, $p < 0.01$). Participants were expected to only fill in the survey when they started their studies in 2010, 2011 or 2012. Excluding the students who started their studies before 2010 yielded similar results (see Appendix A-III, Table II and III). For students from 2010, 2011 and 2012, motivation scores were also highest for scenarios C(+) (M=5.4, SD=1.3) and E(+) (M=4.8, SD=1.4) and lowest for scenario E(-) (M= 1.4, SD=0.5). The difference in motivation score between E(+) and E(-) was significant ($F$ (1, 13) = 26.65, $p < 0.01$); this was also the case for C(+) and E(-) ($F$ (1, 8) = 38.10, $p < 0.01$).

Table 1

*Mean motivation scores*

| Scenario | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Topic | + | + | + | + | + | + | + | + |
| Sample size | 10 | 16 | 9 | 17 | 13 | 9 | 12 | 8 |
| Motivation μ±σ | 4.3±1.9 | 4.6±1.7 | **5.2±1.2** | 4.2±5.1 | **5.1±1.4** | 3.8±0.8 | 4.8±1.8 | 4.4±1.5 |
| | | | | | | | | |
| Topic | - | - | - | - | - | - | - | - |
| Sample size | 16 | 10 | 16 | 8 | 9 | 14 | 9 | 13 |
| Motivation μ±σ | 3.6±1.4 | 2.9±1.6 | 4.1±1.8 | 3.5±1.6 | **2.0±1.2** | 3.4±1.9 | 2.7±1.5 | 3.0±1.5 |
| MSMS* | 3.8 | 4.0 | 4.5 | 4.0 | 3.8 | 3.6 | 3.9 | 3.5 |

*Note:* Mean motivation scores μ and standard deviations σ for every scenario and topic combination. Topics that were expected to be interesting to the participants are designated with '+' and topics that were expected to be discouraging are designated with '-'. *Mean Scenario Motivation Score.

### *Discussion and search task selection*

The goal of this pilot study is to select two scenarios that seem to motivate or discourage participants the most so that in the main experiment of this study participants can be grouped and compared by the degree to which they are motivated. With that goal in mind the search task and

topic in scenario C(+) and E(+) with average scores of 5.2 and 5.1 appear to motivate participants most. The least motivating search task and topic is scenario E(-) scoring 2.0. For extra clarity, examples of the scenarios C(+), E(+) and E(-) are translated and shown in Figure 3.

| Scenario C (+) |
| --- |
| For the course 'academic competences' you are expected to inquire about the following topic: "The development of memory from birth until age 4". Write a short essay of 500 words and relate the topic above to an article in the media. |

| Scenario E (+) |
| --- |
| For the course 'academic competences' your skill will be assessed with regard to writing an objective article. You are therefore required to inquire about the topic 'the development of memory from birth until age 4' and write an objective article about this topic. |

| Scenario E (-) |
| --- |
| For the course 'academic competences' your skill will be assessed with regard to writing an objective article. You are therefore required to inquire about the topic 'the political situation of Guatamala in the 18th century' and write an objective article about this topic. |

*Figure 3.* Scenarios C(+), E(+) and E(-) of the pilot study.

The Likert scale was labeled so that the 1st point indicated 'not motivated' and 7th point indicated 'highly motivated'; points in between were not labeled. Although the points in between were not labeled, it can be reasonably assumed that the 2nd point on the scale, which is also the average motivation score of search task E(-), can be regarded as 'barely motivated' or similar. The interpretation of the 5th point, which corresponds with the average motivation scores of the C(+) and E(+), is more difficult, but can be reasonably expected to represent 'fairly motivated' or stronger. The mean difference in motivation score (3.1) between the E(+) and E(-) search tasks is therefore considered large enough to let them serve *as a basis* for the experimental conditions 'minimal motivation' and 'maximal motivation' in the main experiment of this study. The search tasks will be modified for the main experiment to further manipulate and increase the gap in motivation in order to be as minimally motivating or maximally motivating as possible.

The main reason to select search task E(+) over search task C(+) is that the search task type in scenario E(+) is the same as in scenario E(-). The search task type can therefore not act as a confounding variable in the main experiment of this study. This advantage outweighs the minimal (0.1) decrease in expected average motivation.

**Procedure Determination**

***Screen capture video software.*** When participants perform the search task that was selected in the previous section during the experiment in this study, it will be possible to record their search actions using screen capture video software. Although this captures the actions of the participants, it does not 'record' the credibility evaluations or credibility concerns of the participant i.e. their cognitive activities. These need to be elicited for which think-aloud methods are suitable (Van Someren, Barnard, & Sandberg, 1994). Performing the think-aloud method, participants verbalize what they are thinking, which in this study includes possible credibility evaluations and/or concerns. Think-aloud methods can be employed during the search task or afterwards, which is then called retrospective think-aloud.

***Action protocol - retrospective think-aloud method.*** In this study the 'action protocol - retrospection' method (Van Someren et al., 1994) is chosen in which an action protocol serves as the visual cue to support retrospective think-aloud. The action protocol in this case is the screen capture video that will be recorded during search task execution. Although a 'normal' think-aloud method generally does not interfere with the participants' thought processes related to the task at hand (Van Someren et al., 1994), the retrospective think-aloud method is preferred in order to completely eliminate this possibility. Another reason to prefer retrospective think-aloud over 'normal' think-aloud is the nature of the thought processes that are relevant for this study. The participant might not be aware of for instance heuristic credibility evaluation processes due to their automated nature. The retrospective think-aloud method with visual cues from the screen recorded search task allows more time for the participant to recall and reflect on their thought processes, and are therefore expected to be more suitable to elicit possible credibility evaluations. This is further enhanced by giving the participants the option to pause the recorded screen capture video, if they feel the need to elaborate on their thought processes that occurred during the search task.

Participants performing the retrospective think-aloud method can possibly 'forget' what they were thinking during search task execution. Several measures are therefore taken to

minimize the chance for this possible drawback to occur. First, the screen-capture video that is made during task execution will be played back to the participant to serve as a visual cue for his actions and thoughts when performing the retrospective think-aloud method. Second, the screen-capture video will be played back directly after performing the search. Third, the retrospective think-aloud method will be performed without time limit so that participants have ample time to remember what they were doing. Fourth, participants are allowed to pause their screen capture video during playback when they feel they need more time to elaborate or remember particular items. The researcher will be present during search task execution and during the retrospective think-aloud exercise. This allows the researcher to write down questions that can be asked during the semi-structured interview that follows next.

***Semi-structured Interview.*** The semi-structured interview takes place to allow the researcher to let the participant clear up or elaborate on specific verbalizations and specific actions so that the data is unambiguous and can be coded properly during data analysis.

**Data Analysis Selection**

The goal of this study is to explore what the role of motivation is in credibility evaluations of online information within the entire information seeking process. The model of Metzger (2007) as seen in Figure 2 is limited to credibility evaluations related to a single webpage. It is therefore unclear if the model of Metzger can be applied to the entire information searching process. The literature review also showed that Metzger's model is primarily based on theoretical notions and has only been tentatively supported by empirical research. Instead of analyzing to what degree the model of Metzger is supported by the data, a new framework will therefore be developed through coding the verbatim of the retrospective think-aloud protocols.

Coding verbatim can follow the traditional grounded theory approach through open-coding, axial-coding and selective coding (Glaser & Strauss, 1967), where it is emphasized to enter the coding phase without any preconceptions about possible data structures and concepts. Alternatively, other researchers (King, 1998; Ritchie, Spencer, & O'Connor, 2003) have emphasized that is difficult to have a completely fresh perspective and have developed methods such as the framework analysis and template analysis in which the researcher is encouraged to use and/or form preconceptions as a starting point for the data analysis. Using the framework analysis developed by Ritchie and Spencer (1994), the researcher first familiarizes himself with the data and identifies major themes and possible relations after which coding the data starts.

***Template analysis.*** Using the template analysis (King, 1998) the starting template (before the actual coding of data) consists of codes and themes based on for instance conceptual models found in literature. Given the literature research performed in this study and the models found with regard to credibility evaluations (i.e. integrated model of trust in online information, HSM, and the dual-processing model of credibility assessment), a template analysis was chosen as the verbatim data analysis method. The initial coding template that will be used for coding the verbatim data will be therefore be based on the literature presented in this thesis.

***Audit trail.*** During the coding process the initial coding template will be adapted by the researcher as he learns more about the underlying themes and concepts. To prevent that the final coding template 'appears out of nowhere', an audit trail (Lincoln & Guba, 1985) will be available in Appendix B, Table IV-VI. The audit trail will include the initial template for coding and the meaning and interpretation for each code. After coding 6 and 12 verbatim protocols of the retrospective think-aloud exercises, intermediate templates will be logged. These intermediate templates will include changes in codes and their interpretations so that the development of the initial template into the final template can be inspected. The final template after coding 15 verbatim protocols will include the final coding scheme, the meaning of the codes, clear definitions what is included/excluded from a particular code, and higher level concepts to which the codes belong. After the template is complete and the different concepts are identified the interviews with the participants that were recorded after their retrospective think-aloud protocols will be used to identify and support the different relations/links between codes and concepts. Together, the template analysis and the interviews should allow a comprehensive view of how motivation influences credibility evaluations in the information searching process.

***Inter-rater agreement: Cohen's kappa.*** The reliability of the final coding template first needs to be determined before the analysis of the codes and quotations can start. Without verifying the reliability of the final coding template, it remains uncertain if codes were assigned in a consistent manner and if the coding template descriptions are unambiguous. Determining the reliability of the final coding template can be achieved by coding a portion of the quotations twice and calculating the inter-rater agreement. Van Someren et al. (1994) suggest that the percentage of agreement between the first coding and second coding can serve as an optimistic estimate of inter-rater agreement, and that Cohen's kappa ($\kappa$) can serve as a conservative estimate of inter-rater agreement. Given that multiple codes are possible for each verbatim segment and

that each verbatim segment can be assigned multiple codes, no single measure of Cohen's kappa can convey the necessary information on reliability when some codes may be reliable and others not (Kraemer, Periyakoil, & Noda, 2004). Percentage of agreement and Cohen's kappa will therefore be calculated for each code separately.

## Method

This section will address how the experimental design considerations discussed in the previous section were carried out during the main experiment.

### Participants

In total 19 psychology students registered for the experiment. Students were told that they were only eligible when they had started their university level education in 2011 or 2012. Four students' data were removed for the following reasons. One student did not meet the eligibility criteria, since she started her psychology study in 2010. Another student indicated to be very familiar with a topic she was not expected to be: "Nations and borders in Central-America during the early 19th century". After 15 minutes performing the search task, she realized and verbalized that she thought she was supposed to look for information on the United States of America. Due to the possibly different level of motivation related to the misinterpretation of the search task, her dataset was removed from further analysis. Two other participants' datasets were excluded from further analysis because the attempted manipulation goal of their condition failed (condition 4: minimal motivation - maximal domain expertise). Details on this failure are explained in the following section 'Task'.

The average age of the remaining 15 participants was 20.5 years (SD=1.2) and consisted of 2 male and 13 female students; 3 started in 2011 and 12 started in 2012; 3 were Dutch, 12 were German. Students were compensated for taking part in the experiment by rewarding them with 1 course credit. Each course credit stands for taking part in a psychology experiment for 1 hour. To complete their first year, students have to collect 15 of these credits.

### Task

The goal of this study is to explore how motivation influences credibility evaluations of online information. To this end the primary goal in the main experiment is to manipulate the motivation of the participants. As indicated in the pilot study, search task type and topic E(+) and E(-) seemed most promising to accomplish this goal and thus served as the basis for the two main

conditions of this experiment. However, if the goal is to explore the influence of motivation on credibility evaluations then it needs to be noted that semantic and surface credibility evaluations are enabled by user characteristics i.e. domain knowledge and information skills respectively (Lucassen & Schraagen, 2011). Manipulating motivation without accounting for these user characteristics might confound the results of this study.

All participants are expected to have high information skills compared to the general population after following the university level course 'academic competences'. Having high information skills should enable all participants to perform surface credibility evaluations if they feel the need to and can thus be excluded as a possibly confounding variable.

Domain expertise might confound the results for instance when comparing motivated participants with high domain expertise to barely motivated participants with barely any domain expertise. If the motivated participants perform semantic credibility evaluations, while the barely motivated participants do not, then it would be wrong to conclude that this difference was caused by the difference in motivation. It simply might not have been possible for participants with barely any domain expertise to perform semantic credibility evaluations. For this reason it was attempted to create a 2 x 2 between subjects design in which one variable was motivation (minimal/maximal) and the other was domain expertise (unfamiliar/very familiar).

***Condition 1 goal: Minimal motivation - Minimal domain expertise.*** The goal of condition 1 is to discourage participants as much as possible, while letting the participants search for a topic that they are unfamiliar with. As indicated in the pilot study, selecting search task E(-) for condition 1 is expected to barely motivate participants (motivation score: M=2.0, SD =1.4). Participants also indicated to be barely familiar (M=1.8, SD=1.7) with the topic 'The political situation in Guatamala in the early 19th century'. However, as far as it can assumed that topic interest in this study is a reliable indicator of participant motivation (Pearson's product-moment correlation coefficient r = 0.74, N=189, p <0,01), there was still room for improvement with regard to lowering the expected motivation of participants. The supposedly uninteresting topic still scored on average 3.0 (SD=1.9) on the topic interest scale. Reasons for being interested in the topic were for instance learning more about the culture in Guatamala. In order to eliminate these possibly motivating aspects, the topic was generalized to Central-America and focused on the generic topic of 'nations and borders' instead of politics.

***Condition 2 goal: Maximal motivation - Minimal domain expertise***. Search task E with the topic 'The development of memory in infants from 0-4 years old' should, according to the pilot study, fairly motivate participants (M=5.1, SD=1.4). Participants of the pilot study that started their education in 2011 or 2012 did, however, rate their familiarity with the topic on average 5.0 (SD=1.8), which can be reasonably assumed to indicate they are at least fairly familiar with the topic. Given that the topic is covered in a 3rd year bachelor course of psychology and that 1st and 2nd year students have not yet followed this course, it can be argued that high scores on topic familiarity do not represent high domain expertise. Having high domain expertise requires you to be very familiar with the topic, but being very familiar with a topic does not, however, require you to have high domain expertise. High topic familiarity will therefore no longer be used as a measure for high domain expertise. The classification of domain expertise will be based on the courses that students have followed. It needs to be noted that the classification of 'minimal domain expertise' can still be inferred from very low measures of topic familiarity, since indications of being barely familiar with a topic while at the same time being a domain expert are paradoxical. In condition 2 participants are- thus -expected to have *minimal domain expertise* since they have not yet followed the relevant 3rd year course, despite the possibility that they indicate to be fairly familiar with the topic.

***Condition 3 goal: Maximal motivation - maximal domain expertise.*** The topic 'risk perception and gambling addiction' that featured in the pilot study, was the default candidate for this condition, since the related topics are addressed in 1st and 2nd year psychology courses. However, several participants indicated not to be familiar with the topic. As a precautionary measure a new topic was selected within the domain of psychology of which they surely would be knowledgeable: 'Changes in personality traits during life'. Psychology students in this experiment are expected to be very knowledgeable about, interested in and motivated by the topic since 'personality traits' are covered in the 1st semester of their psychology study. In order to further motivate participants in condition 2 and 3, they were told that objective article that they were expected to write, would serve as input for other participants in this study; in reality this was not the case.

***Condition 4 goal: Minimal Motivation - Maximal domain expertise.*** The creation of this condition was attempted by selecting the topic 'the relation between the significance and power of a statistical test' and the topic 'nations and borders in Europe in the early 19th century'.

Unfortunately students indicated in both cases not be familiar with the topic, despite following statistics courses or having studied nations and borders in Europe in secondary education. It appears to be a daunting task to find a topic that does not motivate participants, but that they are all very knowledgeable about, while still be complex enough so that participants can spend 20 minutes searching for information on the topic. Since the primary goal of this study is to explore the role of motivation, and since inferences about the possibly confounding effect of 'domain expertise' can still be made by comparing conditions 1,2, and 3, no further attempts were made to generate the fourth condition.

**Apparatus**

Participants performed their search tasks on a desktop computer (Intel Core i5-3570, 3.4 GHz) featuring Windows 7. Participants could choose their favorite browser out of 'Internet Explorer', 'Mozilla Firefox' or 'Google Chrome'. Screen capture videos were recorded using Faststone Capture on a 21" AOC LCD display. Audio was recorded using an ICIDU MI-707209 microphone.

**Design**

Independent variable is the topic that is present in each search task. This resulted in a 3x1 between subjects design. Dependent variables measured on 7 point Likert scales are motivation, topic familiarity and topic interest. The first point on the scale indicated 'not motivated', 'not familiar' and 'not interested' respectively. The seventh point on the scale indicated 'highly motivated', 'highly familiar' and 'highly interested' respectively.

**Procedure**

Students were first asked to sign an informed consent form as shown in Appendix C-I. To see if students met the eligibility criteria students were asked for their age, nationality and when they started their university education. The procedure of the experiment was explained afterwards which can be seen in Appendix C-II, Figure I. Students were then shown the search task depending on experimental condition (shown in Appendix C-III, Table VIII) and asked to shortly write down what they thought about the task. Afterwards students were asked to indicate to which degree they were motivated, familiar with the topic, and interested in the topic on 7 point Likert scales. Students then performed the search task for around 20 minutes while the video screen capture was active. Participants were then shown a two minute think-aloud instruction video. Directly afterwards participants were shown their own screen capture recording of the search task

they had performed and performed the retrospective think-aloud method using the just recorded screen capture as a visual cue. Besides the audio being recorded the retrospective -think aloud was also recorded using screen capture. This allowed the researcher to have one video that contained screen-capture video of the participant combined with the participants think-aloud verbalizations to guide the semi-structured interview that was performed to conclude the experiment.

**Data Analysis**

The first step of the data analysis was the development of the initial coding template. The codes constituting the initial coding template followed from an analysis of the reviewed literature in this thesis.

***Code family: timestamps.*** In order to determine when credibility evaluations take place during the information seeking process, it was required to give each quotation (segment of the verbatim protocol) a 'timestamp' code. For the initial coding template the timestamp codes were based on the different information seeking steps that an information seeker takes when searching for information. Several models have been developed to describe these steps of which a comprehensive overview is given by Hearst (2009). Information seeking models typically describe that the information seeker expresses his problem formulation of the information need as a query in a search engine and examines the results. After examining the results the problem is reformulated and entered as a new search query. The information seeker will iterate this cycle until the information need has been fulfilled. From these steps 5 codes were derived listed in Table 2 as the first 5 codes in the code family 'timestamps'. In addition, 5 codes were added to the timestamp code family since they are also observable actions on screen and are all expected to occur in order to complete the search task e.g. navigating, copy/pasting information.

***Code family: Credibility Evaluations.*** The information seeker's decision to use information that is found is made by the information's relevancy to the search task and by the trust in the reliability of the information (Marchionini & White, 2008). It is therefore to be expected that relevancy will play a key role in the information searching process and that verbalizations about the relevancy of information to the search task are present in the data. It is therefore included in the initial template as ‘evaluating relevancy’. The main interest of this thesis, however, are the evaluations that participants are expected to form when assessing the credibility of online information. 'Credibility evaluations' is therefore identified as a separate theme

consisting of the following codes based on the integrated model of trust (Lucassen, 2013): ‘evaluating source credibility’, ‘evaluating surface credibility’ and ‘evaluating semantic credibility’. Generally speaking, quotations will be coded with 'evaluating source credibility' when participants make verbal inferences about the credibility of a source (i.e. the producer of information). 'Evaluating surface features' will be assigned to quotations when participants use information features of an information source as indicators of credibility. 'Evaluating semantic credibility' will be assigned to codes when participants use their knowledge of the topic at hand (i.e. their domain expertise) to determine the credibility of information present. Additionally, ‘raising credibility concerns’ was added as a code for cases when participants are concerned with credibility without directly examining an information source or when it is unclear which type of credibility evaluation was performed. The heuristic-systematic processing distinction was not used in the initial coding template. All resulting codes and corresponding quotations in the credibility evaluation code family will be analyzed in the final coding template to see if a heuristic-systematic classification is warranted, or that a different classification is more suitable.

Table 2

*Initial Coding Template*

| Timestamps | Credibility Evaluations | Information Searching Skills | Other |
|---|---|---|---|
| Direct URL | Evaluating Semantic Credibility | | Sensemaking |
| Search Engine Selection | Evaluating Surface Credibility | | Uncoded |
| Search Entry | Evaluating Source Credibility | | Evaluating Relevancy |
| Examining Search Results | Raising Credibility Concerns | | |
| Examining Source | | | |
| Copy/Pasting | | | |
| Working On Document | | | |
| Saving Information | | | |
| Reading Assignment | | | |
| Navigating | | | |

*Note:* Full descriptions and explanations of each code are given in Appendix B, Table IV.

***Code family: Information Searching Skills.*** Participants are expected to use particular information searching skills to find information. For example, participants might use hotkeys 'CTRL+F' to quickly find information on a website. Alternatively, participants might scan section headings and infer the information content instead of reading entire passages. This code family will be open-coded as no specific literature study was performed on this topic. Open-coding means that there are no predetermined codes and that codes will be developed based on the verbalizations of participants that fit the code family description.

***Code family: Other.*** Several codes that were predetermined, but that did not appear to belong to a particular code family were grouped under the code family 'Other'. 'Evaluating Relevancy' was grouped here for this reason.

The timestamps as indicated previously constitute the information retrieval loop. Hearst (2009) indicates that another loop can be identified during the information seeking process: the sense making loop. During the information seeking process participants are expected to think and learn about the search task at hand; corresponding verbalizations will therefore be coded as 'sense making'. It can also be expected that the verbalizations of participants are only descriptive of the an action that can be witnessed in the video. Since these add no value to understanding the thought processes of the participant, they will be coded as ‘uncoded’. Short and meaningless utterances are also coded as ‘uncoded’ for the same reason.

## Results

### Motivation score results

Before each participant completed the information search task, retrospective think-aloud exercise and interview, they were asked to indicate their motivation, topic familiarity and topic interest on the same 7 point Likert scales that were used in the pilot study. Scores for each participant on these scales are listed in Table 3 and are separated by the goal of the experimental condition the participants were in.

Three participants were removed from further study. Participant 8 was removed from condition goal 1 'Minimal motivation - Minimal domain expertise' as she indicated to be very motivated (6). Participant 4 in the condition goal 2 'Maximal motivation - Minimal domain expertise' Topic (unf.)' was removed as she indicated to be fairly familiar (5) with the topic. Although it was previously discussed that fairly high scores of topic familiarity are not

necessarily indications of high domain expertise, participant 4 was still removed as a precautionary measure. Participant 12 was removed from condition goal 3 'Maximum motivation - Maximal domain expertise' as she indicated to only be moderately motivated (3).

A Pearson product-moment correlation coefficient was computed to assess the relationship between the motivation scores of participants and their interest in the topic ($r = 0.81$ , $n=15$ , $p < 0,01$), indicating a strong correlation between motivation and topic interest corresponding with the results of the pilot-test.

Table 3.

*Participant scores separated by condition goal*

| Condition goal | 1 Minimal Motivation Minimal Dom. Exp. | | | | | 2 Maximal Motivation Minimal Dom. Exp. | | | | | 3 Maximal Motivation Maximal Dom. Exp. | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant # | 6 | **7** | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 11 | 12 | 13 | 14 | 15 |
| Motivation | 2 | 3 | 6 | 4 | 2 | 6 | 5 | 5 | 6 | 5 | 5 | 3 | 7 | 5 | 5 |
| Topic familiarity | 1 | 1 | 3 | 2 | 1 | 3 | 1 | 3 | 5 | 4 | 5 | 3 | 6 | 4 | 5 |
| Topic interest | 2 | 2 | 4 | 2 | 2 | 6 | 6 | 6 | 6 | 6 | 7 | 4 | 7 | 5 | 6 |

*Note:* 'Dom. Exp.' stand for domain expertise. Motivation scores for participants 8 and 12 were not misclassified.

**Template Analysis**

The final coding template (as shown in Appendix B, Table VI) includes the descriptions of codes, examples of segments coded, and higher level concepts. Figure 4 represents code prevalence graphically using the final coding template. The center circle shows the three top level themes (i.e. 'usability evaluation', 'information searching skills' and 'other activities') to which the concepts in the middle circle and codes in the outer circle belong. Numbers inside the circle indicate the theme and concept code prevalence *percentage*. Numbers outside the circle indicate the *absolute* prevalence of each code. In total, 38 'credibility evaluation' related codes comprised 7.5% of all codes (excluding codes: 'timestamps' and 'uncoded') and were assigned to 35 quotations.

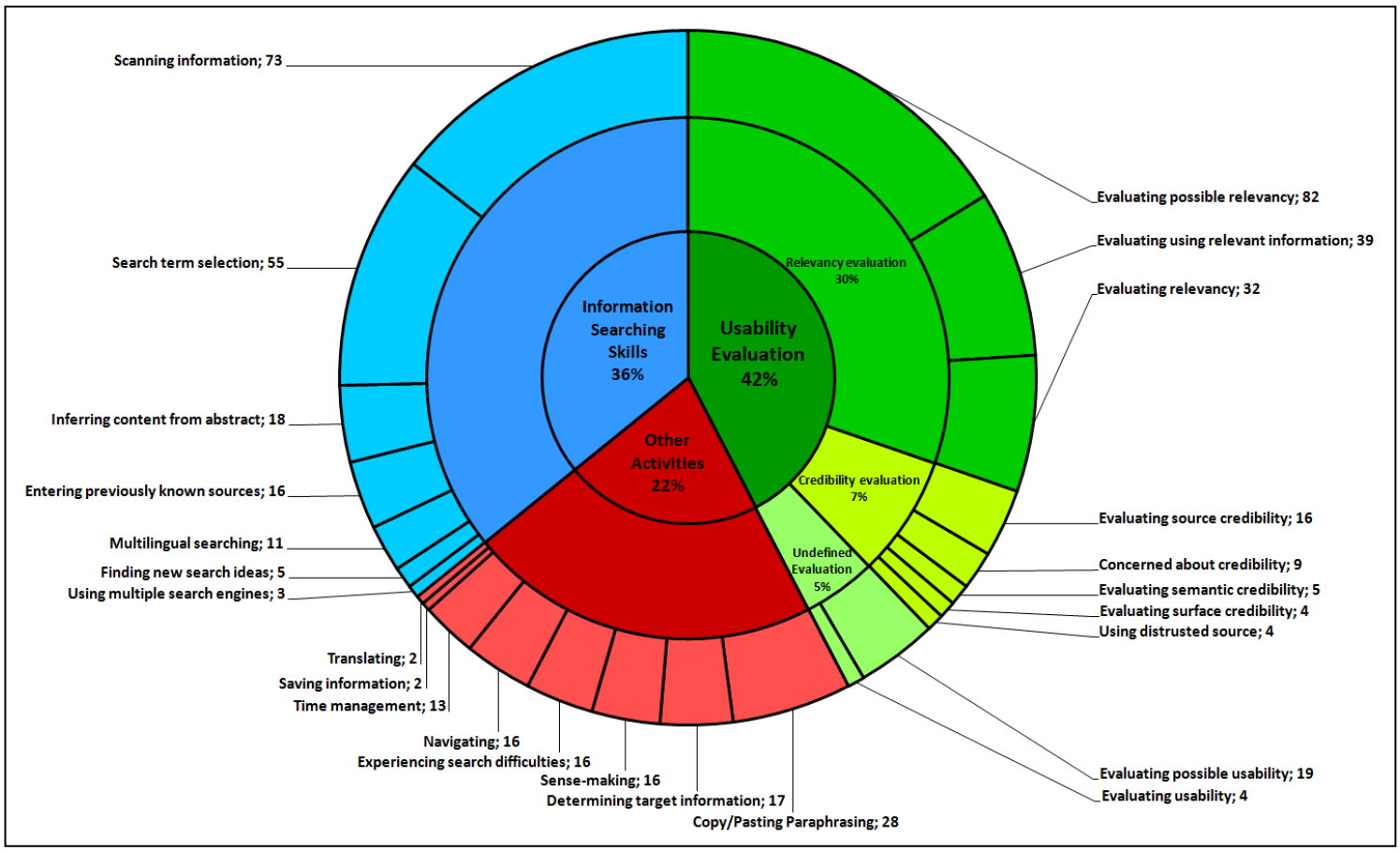


*Figure 4.* Code prevalence representation using the final coding template.

## Coding reliability

Three verbatim protocols (one from each experimental condition) constituting 18.5% of the total amount of quotations were coded twice with three weeks time in between coding sessions to determine coding reliability. Appendix D lists percentage of agreement and Cohen's kappa for all concepts and codes in the final coding template. Inter-rater agreement values related to credibility evaluations are reproduced here, in Table 4.

As stated previously 'percentage of agreement' is considered an optimistic estimate of inter-rater agreement and Cohen's kappa can be considered an conservative estimate. Given that low prevalence of a code to be identified can result in decreased values of kappa, coding reliability regarding credibility evaluations is classified to be at least 'substantial', 'good' or 'sufficient' when using classifications by Landis and Koch (1977); Altman (1991); Fleiss, Levin

and Paik (2003); Van Someren et al. (1994) respectively. The same classifications can be assigned to codes in the categories 'Other' and 'Information Searching Skills'.

Codes related to 'evaluating relevancy' or 'undefined evaluation' can be assigned 'fair', 'poor' and 'insufficient' kappa ratings by the same classification (except for the code 'evaluating using relevant information').

Table 4

*Inter-rater agreement values for credibility evaluations*

| | | CP% | PoA | Cohen's κ |
|---|---|---|---|---|
| Credibility evaluation | | 17 | 0,95 | 0,84 |
| | Evaluating source credibility | 6 | 0,96 | 0,71 |
| | Concerned about credibility | 5 | 0,99 | 0,90 |
| | Evaluating semantic credibility | 1 | 0,99 | 0,66 |
| | Evaluating surface credibility | 4 | 0,97 | 0,65 |
| | Using distrusted source | 1 | 0,99 | 0,66 |

*Note:* CP% = Code Prevalence % in first coding session, and PoA = Percentage of Agreement.

**Credibility evaluations by 'timestamp'**

All credibility evaluation related codes occurred during three different steps or 'time stamps' in the information seeking process (see Table 5): URL Entry - Search Engine Selection (2x), Examining Search Results (9x), Examining Source (23x).

Table 5

*Credibility evaluation related codes by 'timestamp'*

| Timestamp / Credibility code | URL Entry - Search Engine Selection | Examining Search Results | Examining Source |
|---|---|---|---|
| Using distrusted source | - | 1 | 3 |
| Concerned about credibility | 2 | 1 | 4 |
| Source credibility evaluation | - | 5 | 9 |
| Semantic credibility evaluation | - | 2 | 3 |
| Surface credibility evaluation | - | - | 4 |
| Total | 2 | 9 | 23 |

The two quotations related coded 'concerned about credibility' during 'URL Entry - search engine selection' as seen in Table 5 are reproduced here (and translated) to further demonstrate that participants were not only concerned with credibility at the presence of new information, but also during search engine selection:

(1) "I went online and mostly searched using Scopus, since it returns scientific articles. I also looked at Scholar, since it is easier there to find more information." (Participant 1)

(2) "Then I went to the browser and on Google Scholar to find scientific articles on the topic: changes in personality." (Participant 14)

Beside the verbalizations during the think-aloud exercise, participants mentioned similar credibility concerns during the semi-structured interview when discussing search engine selection. All quotations during the interviews regarding search engines are therefore listed in Appendix E.

**Credibility evaluations by experimental condition**

The main differences in credibility evaluation related codes based on experimental condition are given in Table 6.

Table 6

*Credibility evaluation code family - prevalence by condition goal*

| Conditition goal / Credibility code | 1 Minimal Motivation Minimal Dom. Exp. | 2 Maximal Motivation Minimal Dom. Exp. | 3 Maximal Motivation Maximal Dom. Exp. |
|---|---|---|---|
| Using distrusted source | 2 (2x Wikipedia) | 1 (1x Wikipedia) | 1 (1x Wikipedia) |
| Concerned about credibility | 1 | 5 | 3 |
| Source credibility evaluation | 7 (6x Wikipedia) | 4 (1x Wikipedia) | 5 (1x Wikipedia) |
| Semantic credibility evaluation | - | - | 5 (1x Wikipedia) |
| Surface credibility evaluation | - | 4 | - |
| Total | 10 (8x Wikipedia) | 14 (2x Wikipedia) | 14 (3x Wikipedia) |

*Note*: Code prevalence based on experimental condition. Numbers within brackets indicate the amount of quotations in that cell concerning Wikipedia.

It needs to be noted that 5 quotations coded with 'Source credibility evaluation' in condition goal 1 'Minimal motivation - Minimal domain expertise' regarding Wikipedia were repeated utterances by the same participant. All quotations (as shown in Appendix F, Table IX) stated more or less the same short sentence "I only see Wikipedia page which are not very reliable". This was also reflected in her interview where she states her thoughts towards using Wikipedia:

> "Because there [Wikipedia] anyone can write something down so I can't be sure if the information is correct (...) so I would rather not use it." (Participant 9).

No such repeated statements on one information source were found in credibility evaluation related quotations in Appendix F. However, it needs to be noted that all surface credibility evaluations were performed by one participant in condition goal: 'Maximal motivation - Minimal domain expertise'.

## Discussion

This section will first interpret and evaluate the findings as presented in the 'Results' section. This is followed by a general discussion where the experiment design and theoretical framework is evaluated in light of the results. Based on this evaluation some new ideas for future research are presented.

### Discussion of results

#### Motivation score results

Table 3 presented motivation scores from participants after they had read the search task. The goal of condition 1 was to minimally motivate participants. Interpreting the results, it can reasonably assumed that 'minimal motivation' in this case resulted in participants being 'barely/moderately' motivated given their motivation scores (2,3,4,2) and the fact that a score of 1 represents 'no motivation'. Following a similar analysis it can be reasonably assumed that the goal of maximally motivation participants in condition 2 and 3 resulted in participants being 'fairly/strongly' motivated, given their motivation scores (6,5,5,5,5,7,5,5) and the fact that a score of 7 represents 'highly motivated'.

It can be questioned whether listed motivation scores represent actual motivation of participants. This can only pose a problem for drawing conclusions based on experimental conditions, when low motivation scores are in reality higher, and/or when high motivation scores are in reality lower.

Topic interest scores of participants ('minimal motivation' condition = 2,2,2,2; maximal motivation condition = 6,6,6,6,7,7,5,6) and can serve as an indicator for possibly inflated or deflated motivation scores given the strong high correlation of topic interest with motivation score (Pilot study Pearson's r = 0,74; Main experiment r = 0,81). Topic interest scores in this regard, would suggest that low motivation scores are possibly inflated, and high motivation scores are possibly deflated strengthening the idea that the manipulation through experimental condition was successful. Additionally, first reactions to the search task in each condition were written down by participants (as seen in Appendix C-III) and visually inspected to see if they matched descriptions of participants in the pilot study who indicated to be motivated or discouraged. The only reaction that stood out was of participant 9 who, besides indicating that she did not like the topic and that she was not knowledgeable about the topic, also gave her positive impressions of the research topic of this thesis. Her positive impressions of the thesis topic might have inflated her motivation score (4). Because of this additional support for a successful manipulation of motivation, participants in the 'minimal motivation' condition will be referred to as being 'barely motivated' and participants in the 'maximal motivation' condition will be referred to as being 'strongly motivated'.

**Template analysis results**

Besides the fact that code prevalence is not necessarily proportional to the amount of time spent on a certain activity nor an indication of the nature of each coded segment, Figure 4 seems to indicate that participants devote relatively large amounts of effort evaluating the relevancy of information (30% of code prevalence), employing several information skills to quickly find information ('scanning for information' code prevalence = 14%). Conversely, credibility evaluation related codes only comprised 7% of the total amount of codes assigned. These numbers seem to indicate that a participants' primary concern is to find relevant information. This is not surprising, as it is generally better to find relevant information with unsure reliability than reliable information that is not relevant to the search task at hand. Appendix G further illustrates this point by listing which codes occurred during each 'timestamp'.

**Coding reliability**

Although coding reliability was classified as 'substantial/good/sufficient' for credibility evaluation related codes, this was not the case for relevancy evaluation related codes and usability evaluation related codes. The most likely reason for this is that 'Evaluating possible

usability' and 'Evaluating usability' were added to the coding template during the last development step in an attempt to classify 'undefined usability evaluations' separately. When usability evaluation and relevancy related codes are grouped together (excluding 'evaluating using relevant information'), the percentage of agreement between coding sessions becomes 0,85 and κ becomes 0,64 indicating 'substantial/ good/sufficient' coding reliability. Future use of these coding categories thus warrants better explanations and descriptions of each codes.

**Credibility evaluations by experimental condition**

Discounting the repeated statements of participant 9 with regard to source credibility evaluation, it becomes very clear that barely motivated participants were hardly concerned with credibility. This is further illustrated after investigating the further nature of credibility evaluation quotations by 'barely motivated' participants: 8 out of 10 credibility related codes regarded Wikipedia. The fact that low motivation participants hardly performed any credibility evaluations and showed negligible credibility concern resembles the prediction made by Metzger's (2007) model i.e. not properly motivated information seekers will not perform meaningful credibility evaluations. This as opposed to 'strongly motivated' participants who did perform meaningful credibility evaluations as shown in Appendix F Table X and XI.

The semantic and surface credibility evaluations performed by 'strongly motivated' participants is indicative of the enabling character of personal characteristics with regard to credibility evaluations as presented in the trust in online information model developed by Lucassen (2013). Semantic credibility evaluation require that the participant has domain expertise. It is therefore not surprising that participants who performed semantic credibility evaluations were 'strongly motivated' and considered domain experts as they had just followed a university level course related to the search task topic.

All participants were expected to have similar amounts of information skills as discussed during participant selection. The fact that surface credibility evaluations were only performed by 'strongly motivated' participants further supports Metzger's model that meaningful credibility evaluations are only performed when properly motivated and enough ability (in this case: information skills).

So far, all findings seem to correspond with Metzger's dual-processing model of credibility assessment and Lucassen's integrated model of trust in online information. However, Metzger's dual-processing model of credibility assessment also predicts that systematic

processing occurs when information seekers are motivated and have enough ability. No systematic processing was detected in any of the credibility evaluations presented in Appendix F.

**Credibility evaluations by 'timestamp'**

An explanation for the lack of systematic processing with regard to credibility evaluations and concerns, might be that participants have developed a heuristic that guarantees reliable information and therefore negates the need for systematic processing. There are several findings in the codes and quotations that support this idea and that the heuristic is: 'selecting a search engine e.g. scholar.google.com, sciencedirect.com, that only returns search results that were published in reliable information sources such as books and journals'. Three separate findings support this notion.

*First,* on two instances participants raised credibility concerns during the search engine selection process as stated in Table 5:

(1) "I went online and mostly searched using Scopus, since it returns scientific articles. I also looked at Scholar, since it is easier there to find more information." (Participant 1)

(2) "Then I went to the browser and on Google Scholar to find scientific articles on the topic: changes in personality." (Participant 14)

*Second*, on several other occasions did 'strongly motivated' participants condition select scientific search engines without verbalizing anything and were therefore not coded as being concerned about the credibility of the information. Investigating the interviews that were taken after the retrospective think-aloud exercise with regard to scientific search engine selection indicates that selecting scientific search engines is largely done specifically to find only scientific/reliable/trustworthy/credible information as seen in Appendix E.

*Third,* 'strongly motivated' participants (# 1,2,13,14) *did not perform any credibility evaluations or any credibility concerns as soon as their search in scientific search engines started.* When these participants 'downgraded' to search engines such as Google, where reliable and unreliable information are interspersed, credibility concerns were uttered almost instantaneously:

(1) "Although I do not think Wikipedia is very trustworthy, it can be used for very general statements" (Participant 13)

(2) "No scientifically published articles, but still articles that I can use to perform this assignemnt" (Participant 14)

It seems that within the confines of a reliable search environment students are 'certain' that the information is reliable and that therefore credibility evaluations or concerns are no longer an issue. This is in accord with a statement made by Lucassen (2013):

"(...) trust involves taking a certain risk. Full systematic processing would mean that every aspect relevant for credibility is evaluated, reducing this risk to zero. This means that trust is replaced by certainty. Hence, credibility evaluation is always heuristic to a certain extent." (p.13)

**General discussion**

The results presented indicate that motivation is an essential driver for credibility evaluations and concerns. On the one hand, 'barely motivated' participants were hardly concerned with credibility and did not perform notable credibility evaluations. On the other hand, 'strongly motivated' participants showed considerable credibility concerns and performed heuristic credibility evaluations. Systematic credibility evaluations were not present in this study; instead it appears that participants used another heuristic (using scientific search engines) to replace the need for systematic credibility evaluations.

Although these findings contradict the prediction by Metzger's model (2007) in which participants that are motivated and able should exhibit systematic processing, they are still in accord with the original HSM (Chaiken, 1980) on which Metzger's model is based. This is due to the fact that the HSM emphasizes the value of the 'principle of least cognitive effort' and the sufficiency principle'. The HSM predicts that information seeker's will use heuristic processing as long as it closes the gap between the actual level of confidence and the desired level of confidence. Only when the gap cannot be met by heuristic processing, will the information seeker ensue in systematic processing. The results in this study exemplify this 'principle of least effort' and 'the sufficiency principle', where the information seeker selects a heuristic (scientific search engine selection) in order to make heuristic and systematic credibility evaluations obsolete.

Although the data seem to fit the original HSM, it can be argued that the set-up of the experiment is the cause for the absence of systematic processing. First, it can be argued that the participants in this study were not 'able' enough to perform systematic semantic credibility evaluations. This seems unlikely. If university students, who have just followed a course on 'personality traits' and taken a written exam, do not have enough domain expertise to perform systematic credibility evaluations, then it is unlikely that the average person who is not a domain expert in a specific domain will ever perform systematic semantic credibility evaluations.

Second, in a similar fashion it can be argued that participants in this study did not possess enough information skills to be able to perform systematic surface credibility evaluations. Again, it needs to be noted that these participants have enrolled in a university course called 'academic competences' in which these skills are taught. It can be questioned if the average person, who is likely to possess less information skills than these students, will be able to exhibit systematic surface credibility evaluations.

Third, instead of the skill set of the participants not being adequate an argument in the opposite direction can be made. Perhaps the heuristics (excluding the scientific search engine selection heuristic) that the participants used to perform surface and semantic credibility evaluations were sufficient to close the gap between the actual level of confidence and the desired level of confidence. This would indicate that the motivation manipulation performed to raise the desired level of confidence, despite being successful, was not strong enough to elicit systematic processing. It is possible that this is the case. It needs to be noted that this argument is based on two aspects. *The first aspect* is that the desired level of confidence was not raised high enough to elicit systematic processing. As stated in the pilot study, the scenario's are fictional and there is no 'real' consequentiality for the participants. It can thus be argued that the maximum score of 7 on the Likert scale for motivation pales in comparison to the amount of motivation elicited when writing a thesis like this. Some students in the interviews indicated that they would be more motivated if it was 'for real'. On the other hand, some students also indicated that they 'do not settle for less than reliable information'. In addition, it needs to be noted that participants who are interested in a certain topic and search for information about this topic, have in intrinsic incentive to find reliable information so that they do not store incorrect information or incorrectly change their mental construct of a topic that they are interested in. Furthermore, it is noteworthy that the goal of this study was to capture day-to-day information seeking behavior. High stakes, high consequentiality information seeking, are expected to form a minority in the day-to-day information seeking behavior. *The second aspect*, is that participants in this study can be expected to have for instance, as Lucassen (2013) noted in his doctoral thesis, more efficient heuristics with regard to information skills than high-school students. A future research study can account for this possibility in the following way. In order to elicit 'systematic processing' a promising set-up would be to use high school students interested in a certain topic as participants, and giving them an extra incentive to find information about the topic. Thus, raising the desired level of

confidence, while lowering the efficiency and effectiveness of the participants' heuristic processing. This would enlarge the gap between the desired level of confidence and the actual level of confidence while simultaneously having participants with less effective heuristics to close this gap and who are then forced to ensue in systematic processing.

References

Altman, D. G. (1991). *Practical Statistics for Medical Research.* London: Chapman and Hall.

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum, 36*(2), 3-10.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752-766.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-473.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken, & Y. Trope, *Dual-process theories in social psychology* (pp. 73-96). New York: The Guilford Press.

Dutta-Bergman, M. J. (2004). The impact of completeness and web use motivation on the credibility of e-health information. *Journal of Communication, 54*, 253-269.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly, 77*, 515-540.

Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical Methods for Rates & Proportions.* New York: Wiley and Sons.

Fogg, B. J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. *ACM Conference on Human Factors in Computing Systems (CHI'03), extended abstracts* (pp. 722-723). New York: ACM Press.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of web sites? A study with over 2,500 participants. *Proceedings of the 2003 conference on Designing for user experiences (DUX '03)* (pp. 1-15). New York: ACM.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine Publishing Company.

Hearst, M. (2009). *Search User Interfaces.* Cambridge: Cambridge University Press.

Hilligoss, B., & Rieh, S. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management, 44*, 1467-1484.

Kellar, M., Watters, C., & Shepherd, M. (2006). A goal-based classification of web information tasks. *Proceedings of the American Society for Information Science and Technology Conference, 43*, pp. 1-22.

Kraemer, H. C., Periyakoil, V. S., & Noda, A. (2004). Kappa coefficients in medical research. In R. B. D'Agostino, *Tutorials in Biostatistics, Volume 1, Statistical Methods in Clinical Studies* (pp. 85-105). Chichester: John Wiley & Sons.

Kubiszewski, I., Noordewier, T., & Costanza, R. (2011). Perceived credibility of internet encyclopedias. *Computers & Education, 56*, 659-667.

Landis, J., & Koch, G. (1977). Measure of Observer Agreement for Categorical Data. *Biometrics*, 159-174.

Lucassen, T. (2013). *Trust in online information.* Enschede: Ipskamp.

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62*, 1232-1242.

Lucassen, T., Muilwijk, R., Noordzij, M. L., & Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology, 64*, 254-264.

Marchionini, G., & White, R. W. (2008). Find what you need, understand what you find. *Journal of Human-Computer Interaction, 23*, 205-237.

Marchionini, G., & White, R. W. (2008). Find What You Need, Understand What You Find. *Journal of Human-Computer Interaction, 23*, 205-237.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*, 2078-2091.

Rieh, S. Y., & Belkin, N. J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *Proceedings of the 61st Annual Meeting of the American Society for Information Science, 35*, pp. 279-289.

Ritchie, J., & Spencer, L. (1994). Qualitative Data Analysis for Applied Policy. In A. Bryman, & R. G. Burgess, *Analysing Qualitative Data* (pp. 173-194). London: Routledge.

Ritchie, J., Spencer, L., & O'Connor, W. (2003). Carrying out Qualitative Analysis. In J. Ritchie, & J. Lewis, *Qualitative Research Practice: a Guide for Social Science Students and Researchers* (pp. 219-262). London: SAGE Publications.

Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. *Proceedings of the 13th International Conference on World Wide Web (WWW'04)* (pp. 13-19). New York: ACM.

Seybert, H. (2011). *Internet use in households and by individuals in 2011*. Eurostat Statistics In Focus. Retrieved from http://epp.eurostat.ec.europa.eu/.

Shirky, C. (2008). *Here Comes Everybody: How Change Happens when People Come Together.* New York: Penguin Press.

Stanford, J., Tauber, E. R., Fogg, B. J., & Marable, L. (2002). *Experts vs. online consumers: A comparative credibility study of health and finance Web sites.* Consumer WebWatch Research Report.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical guide to modeling cognitive processes.* Amsterdam: Academic Press.

## APPENDIX A-I: Pilot Study Scenarios and Topics

Various scenarios were presented to participants in the pilot study. 'X' can be replaced by each of the four topics listed after the scenarios.

*Scenario A:*

Voor het vak 'academische vaardigheden' krijg je een artikel te lezen over het onderwerp 'X'. Voor het volgende college dien je over dit artikel en de aanverwante theorieën en begrippen een samenvatting te schrijven voor je medestudenten.

*Scenario B:*

Een onderdeel van het vak 'academische vaardigheden' is het verbeteren van de presentatie vaardigheden van de studenten. Daartoe krijg je bij het volgende werkcollege de volgende opdracht:
'Je hebt 1,5 uur de tijd om jezelf in te lezen m.b.t. het onderwerp 'X' en een presentatie hierover voor te bereiden. Daarna presenteer je voor je groepsgenoten, die de opdracht hebben gekregen om feedback te geven op je presentatie.'

*Scenario C:*

Voor het vak 'academische vaardigheden' dien je jezelf te verdiepen in het onderwerp 'X'. Schrijf een kort essay van 500 woorden en leg daarin de link tussen het bovenstaande onderwerp en een artikel in de media.

*Scenario D:*

Om erachter te komen welke natuurlijke houding jij aanneemt in groepsgesprekken zal je tijdens het volgende werkcollege van 'academische vaardigheden' een discussie gaan voeren met medestudenten. Ter voorbereiding hierop wordt je geacht om jezelf te verdiepen in het onderwerp 'X'.

*Scenario E:*

Voor het vak 'academische vaardigheden' wordt je vaardigheid getest in het schrijven van een uiteenzetting. Daartoe dien je voor het volgende college je te verdiepen in het onderwerp 'X' en hierover een uiteenzetting te schrijven.

*Scenario F:*

Voor het vak 'academische vaardigheden' dien je jezelf voor te bereiden op een individueel mondeling examen over het onderwerp 'X'. Tijdens het examen dien je de hoofdlijnen van dit onderwerp aan te stippen. Aan de hand daarvan zal de docent je een aantal vragen stellen.

*Scenario G:*

Voor het vak 'academische vaardigheden' krijg je als opdracht om jezelf te verdiepen in het onderwerp 'X'. Voor het volgende college wordt je geacht via email een vraag in te dienen over de te bestuderen stof. De ingediende vragen zullen klassikaal behandeld.

*Scenario H:*

Voor het vak 'academische vaardigheden' leer je het schrijven van een onderzoeksvoorstel. Als onderwerp is jou toebedeeld 'X'. Bedenk hierbij een onderzoeksvraag , onderzoeksmethode, en de relevante statistische toesten en mail dit voor het volgende college naar de docent.

*Topics:*

Topics that are expected to be motivating to students, labeled as (+).

Topic 1 (+): 'De ontwikkeling van het geheugen van geboorte tot het 4e levensjaar'

Topic 2 (+): 'Risicoperceptie en aanleg voor gokverslaving'

Topics that are expected to be discouraging to students, labeled as (-).

Topic 3 (-): 'Rijstteelt en de economische situatie van de Filipijnen in de 19e eeuw'

Topic 4 (-): 'De politieke situatie van Guatamala in 18e eeuw'

Table I

*Order of scenario and topic appearance per group in the pilot study*

| Group 1 | Group 2 | Group 3 | Group 4 |
|---|---|---|---|
| Scenario A & Topic 1 | Scenario D & Topic 2 | Scenario E & Topic 4 | Scenario F & Topic 3 |
| Scenario B & Topic 3 | Scenario C & Topic 4 | Scenario F & Topic 2 | Scenario H & Topic 4 |
| Scenario C & Topic 2 | Scenario A & Topic 3 | Scenario G & Topic 3 | Scenario E & Topic 1 |
| Scenario D & Topic 4 | Scenario B & Topic 1 | Scenario H & Topic 1 | Scenario G & Topic 2 |

## APPENDIX A-II: Pilot Study Questions

Actual depiction of questions regarding one of the scenarios as presented to the participants:

**Scenario 1:**

**Voor het vak 'academische vaardigheden' krijg je een artikel te lezen m.b.t. het onderwerp 'De ontwikkeling van het geheugen van geboorte tot het 4e levensjaar'.**

**Voor het volgende college dien je over dit artikel en de aanverwante theorieën en begrippen een samenvatting te schrijven voor je medestudenten.**

| | Niet gemotiveerd | | | | | | Zeer gemotiveerd |
|---|---|---|---|---|---|---|---|
| In hoeverre ben je gemotiveerd om met de bovenstaande opdracht aan de slag te gaan? | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

**Waarom heb je voor het bovenstaande antwoord gekozen?** *

**Welke aspecten van het bovenstaande scenario zijn van invloed op jouw motivatie voor de opdracht?** *

| | Niet bekend | | | | | | Zeer bekend |
|---|---|---|---|---|---|---|---|
| In hoeverre ben je bekend met het onderwerp van de opdracht? | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

| | Niet geïnteresseerd | | | | | | Zeer geïnteresseerd |
|---|---|---|---|---|---|---|---|
| In hoeverre ben je geïnteresseerd in het onderwerp van de opdracht? | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

## APPENDIX A-III: Pilot Study Mean Motivation Scores

Mean motivation scores μ and standard deviations σ for every scenario and topic. Topics that were expected to be interesting to the participants are designated with '+' and topics that were expected to be discouraging are designated '-'. MSMS stands for Mean Scenario Motivation Score.

Table II (all pilot study respondents)

*Mean motivation scores per scenario per topic*

| Scenario | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Topic | + | + | + | + | + | + | + | + |
| Sample size | 10 | 16 | 9 | 17 | 13 | 9 | 12 | 8 |
| Motivation μ±σ | 4.3±1.9 | 4.6±1.7 | **5.2±1.2** | 4.2±5.1 | **5.1±1.4** | 3.8±0.8 | 4.8±1.8 | 4.4±1.5 |
| | | | | | | | | |
| Topic | - | - | - | - | - | - | - | - |
| Sample size | 16 | 10 | 16 | 8 | 9 | 14 | 9 | 13 |
| Motivation μ±σ | 3.6±1.4 | 2.9±1.6 | 4.1±1.8 | 3.5±1.6 | **2.0±1.2** | 3.4±1.9 | 2.7±1.5 | 3.0±1.5 |
| MSMS* | 3.8 | 4.0 | 4.5 | 4.0 | 3.8 | 3.6 | 3.9 | 3.5 |

Table III (only pilot study respondents from 2010, 2011 and 2012)

*Mean motivation scores per scenario per topic*

| Scenario | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Topic | + | + | + | + | + | + | + | + |
| Sample size | 6 | 13 | 5 | 14 | 10 | 5 | 10 | 4 |
| Motivation μ±σ | 4.3±2.0 | 4.3±1.8 | **5.4±1.3** | 4.4±1.6 | **4.8±1.4** | 3.8±0.8 | 4.8±2.0 | 3.5±1.7 |
| | | | | | | | | |
| Topic | - | - | - | - | - | - | - | - |
| Sample size | 13 | 6 | 13 | 4 | 5 | 10 | 5 | 10 |
| Motivation μ±σ | 3.6±1.3 | 3.2±1.8 | 4.1±1.7 | 3.8±1.0 | **1.4±0.5** | 2.6±1.4 | 2.2±1.6 | 2.5±1.0 |
| MSMS* | 3.8 | 4.0 | 4.4 | 4.2 | 3.7 | 3.0 | 3.9 | 2.8 |

## APPENDIX B: Audit Trail - Coding Template Development

Table IV

*Initial Coding Template as used during coding*

| Initial Coding Template | Constitutes: |
|---|---|
| | Thoughts and actions related to: |
| Code Family - Search Chronology | |
| @ Direct URL | Selecting a direct URL other than Search engines to find information |
| @ Search Engine Selection | Selecting a search engine to find target information |
| @ Search Entry | Filling in search queries |
| @ Examining Search Results | Starts directly after '@ Search Entry' and finishes when '@ Examing Source' starts. |
| @ Examining Source | Either through direct URL or through a hyperlink found elsewere (predominantly @ Examining Search Results) |
| @ Copy/Pasting Paraphrasing | Copying and pasting information from a website and later paraphrasing of that content. |
| @ Working On Document | Editing the layout, style, deleting content from the working document |
| @ Saving Information | Saving pdfs or images for later use. (similar to copy pasting) |
| @ Reading Assignment | Obvious |
| @ Navigating | Navigating documents, browsers etc. |
| | |
| Code Family - Credibility | |
| | |
| Raising Credibility Concerns | Whenever participant voices concerns over credibility while not evaluating content or so directly. |
| Evaluating Semantic Credibility | Lucassen |
| Evaluating Source Credbility | Lucassen |
| Evaluating Surface Credibility | Lucassen |
| | |
| Code Family - Information Searching Skills | Skills, tricks or methods employed to find the target information |
| Open Coding | |
| | |
| Separate Codes: | |
| Evaluating Relevancy | Evaluating whether content is relevant to the search task at hand |
| Sense-making | Making sense of the search task and what it all means. |
| Uncoded | - Verbalizations that are solely descriptive of an observable action |
| | - Verbalizations to short or to meaningless to code |

Table V

*Coding template after coding 6 verbatim transcripts*

| Initial Coding Template | Constitutes: |
|---|---|
| | Thoughts and actions related to: |
| Code Family - Search Chronology | |
| @ Direct URL | Selecting a direct URL other than Search engines to find information |
| @ Search Engine Selection | Selecting a search engine to find target information |
| @ Search Entry | Filling in search queries |
| @ Examining Search Results | Starts directly after '@ Search Entry' and finishes when '@ Examing Source' starts. |
| @ Examining Source | Either through direct URL or through a hyperlink found elsewere (predominantly @ Examining Search Results) |
| @ Copy/Pasting Paraphrasing | Copying and pasting information from a website and later paraphrasing of that content. |
| @ Working On Document | Editing the layout, style, deleting content from the working document |
| @ Saving Information | Saving pdfs or images for later use. (similar to copy pasting) |
| @ Reading Assignment | Obvious |
| @ Navigating | Navigating documents, browsers etc. |
| @ Examining Abstract/Introduction | |
| | |
| Code Family - Credibility | |
| | |
| Raising Credibility Concerns | Whenever participant voices concerns over credibility while not evaluating content or so directly. |
| Evaluating Semantic Credibility | Lucassen |
| Evaluating Source Credbility | Lucassen |
| Evaluating Surface Credibility | Lucassen |
| Using distrusted source | |
| | |
| Code Family - Information Searching Skills | Skills, tricks or methods employed to find the target information |
| Open Coding | |
| Finding new search ideas | |
| Inferring content from abstract | |
| Multilingual Searching | |
| Scanning Information | |
| Search Term Selection | |
| Using Multiple search engines | |
| Using Sideline Comments | |
| Using Previously Known Sources | |
| | |
| Relevancy Familiy: | |
| Evaluating Possible Relevancy | Evaluating if information that is yet to be found in the source instead of the abstract, might be relevant |
| Evaluating Relevancy | Evaluating whether content is relevant to the search task at hand |
| Evaluating Usability of Relevant Information | Evaluating if relevant information to the search task/topic is usable for the search task. |
| | |
| Determining Target Information | Deciding what to look for (perhaps result of sensemaking?) |
| Sense-making | Learning about the topic and search task at hand |
| Evaluating Usability Feasibility | Given the time limitations, can I do this or that. |
| Experiencing search difficulties | - |
| Experiencing stress | - |
| Translating | (perhaps part of sense-making?) |
| Uncoded | - Verbalizations that are solely descriptive of an observable action |
| | - Verbalizations to short or to meaningless to code |

Table VI

*Coding template after coding 12 verbatim transcripts.*

*Coding template was not changed after coding 15 verbatim transcripts.*

| Initial Coding Template | Constitutes: |
|---|---|
| | Thoughts and actions related to: |
| Code Family - Search Chronology | |
| | |
| @ Search Engine Selection | Selecting a search engine to find target information |
| @ Search Entry | Filling in search queries |
| @ Examining Search Results | Starts directly after '@ Search Entry' and finishes when '@ Examing Source' starts. |
| @ Examining Source | Either through direct URL or through a hyperlink found elsewere (predominantly @ Examining Search Results) |
| @ Copy/Pasting Paraphrasing | Copying and pasting information from a website and later paraphrasing of that content. |
| @ Working On Document | Editing the layout, style, deleting content from the working document |
| @ Saving Information | Saving pdfs or images for later use. (similar to copy pasting) |
| @ Reading Assignment | Obvious |
| @ Navigating | Navigating documents, browsers etc. |
| @ Examining Abstract/Introduction | |
| | |
| Code Family - Credibility | |
| | |
| Raising Credibility Concerns | Whenever participant voices concerns over credibility while not evaluating content or so directly. |
| Evaluating Semantic Credibility | Lucassen |
| Evaluating Source Credbility | Lucassen |
| Evaluating Surface Credibility | Lucassen |
| Using distrusted source | |
| | |
| Code Family - Information Searching Skills | Skills, tricks or methods employed to find the target information |
| Open Coding | |
| Finding new search ideas | |
| Inferring content from abstract | |
| Multilingual Searching | |
| Scanning Information | Scanning includes searching for predetermined target information, checking headlines, section headings |
| Search Term Selection | |
| Using Multiple search engines | |
| Using Sideline Comments | |
| Using Previously Known Sources | |
| | |
| Usability Family | |
| Evaluating Possible Relevancy | Evaluating if information that is yet to be found in the source instead of the abstract, might be relevant |
| Evaluating Relevancy | Evaluating whether content is relevant to the search task at hand |
| Evaluating Usability of Relevant Information | Evaluating if relevant information to the search task/topic is usable for the search task. |
| Evaluating Possible Usability | When glancing at snippets of the total source and evaluating that is looks or does not look usable/interesting/good but not stating why exactly |
| Evaluating Usability | When glancing at the source and evaluating that is looks or does not look usable/interesting/good but not stating why exactly |
| | |
| When evaluating information users evaluate for usability comprising | |
| Relevancy | |
| Credibility | |
| | |
| Determining Target Information | Deciding what to look for (perhaps result of sensemaking?) |
| Sense-making | Learning about the topic and search task at hand |
| Time Management | Given the time limitations, can I do this or that. + Being aware of time limitations |
| Experiencing search difficulties | - |
| Experiencing stress | - |
| Translating | (perhaps part of sense-making?) |
| Uncoded | - Verbalizations that are solely descriptive of an observable action |
| | - Verbalizations to short or to meaningless to code |
| | |
| | |
| | |
| Prevalent Coding Combinations | |
| It becomes clear that participants are first and foremost involved with finding relevant information. | |
| They employ several strategies to make this process smooth, quick, and effective. | |
| | |
| @ Examining Search Results | |
| The scan for possible relevancy of the short texts | |
| | |
| @ Examining Abstract/Introduction | |
| They scan for possible relevancy | |
| They infer the content of the source | |
| | |
| @ Examining Source | |
| Scanning for information | |
| Evaluating Possible relevancy | |
| | |
| @ Examining Source | |
| Evaluating Relevancy | |
| Sometimes leading to usability of relevant information. | |

## APPENDIX C-I: Informed Consent

Informed consent form as shown to all participants prior to taking part in the experiment.

**GEÏNFORMEERDE TOESTEMMING** GW.07.130

Ik, …….. *(naam proefpersoon)*

Stem toe mee te doen aan een onderzoek dat uitgevoerd wordt door

Ewald Maas

Ik ben me ervan bewust dat deelname aan dit onderzoek geheel vrijwillig is. Ik kan mijn medewerking op elk tijdstip stopzetten en de gegevens verkregen uit dit onderzoek terugkrijgen, laten verwijderen uit de database, of laten vernietigen.

*De volgende punten zijn aan mij uitgelegd:*

1. Het doel van dit onderzoek is de relatie tussen motivatie en het inschatten van de betrouwbaarheid van online informatie te bestuderen.
   Deelname aan dit onderzoek zal meer inzicht geven in de relatie tussen motivatie en het inschatten van de betrouwbaarheid van online informatie.
2. Er zal mij gevraagd worden een studie-gerelateerde opdracht uit te voeren en deze na afloop met de onderzoeker te bespreken.
   Het hele onderzoek zal ongeveer 60 minuten duren. Aan het einde van het onderzoek zal de onderzoeker uitleggen waar het onderzoek over ging.
3. Er behoort geen stress of ongemak voort te vloeien uit deelname aan dit onderzoek.
4. De gegevens verkregen uit dit onderzoek zullen anoniem verwerkt worden en kunnen daarom niet bekend gemaakt worden op een individueel identificeerbare manier.
5. De onderzoeker zal alle verdere vragen over dit onderzoek beantwoorden, nu of gedurende het verdere verloop van het onderzoek.

Handtekening onderzoeker: …… Datum: 18-2-2013

Handtekening proefpersoon: …… Datum: ……..

## APPENDIX C-II: Experiment Procedure

WELKOM!

Hieronder volgt stapsgewijs wat we gaan doen tijdens de komende 60 min. en hoeveel tijd elk onderdeel ong. in beslag zal nemen.

De procedure van het experiment:

1. (5 min.) Uitleg en instructie (dit A4 lezen, ondertekenen informed consent, en uitreiken opdracht).
2. (20 min.) Uitvoeren v.d. opdracht. Activiteiten op de computer worden opgenomen middels screen-capture.
3. (5 min.) Instructie: retro-spective think-aloud.
4. (20 min.) Uitvoeren: retro-spective think-aloud m.b.t. zojuist uitgevoerde opdracht.
5. (10 min) Beantwoorden van vragen van de onderzoeker m.b.t. think-aloud.

*Figure I.* Explanation of the experiment procedure as shown to participants.

## APPENDIX C-III: Search Tasks by Experimental Condition

Search tasks as presented to participants for each condition in the main experiment of this study.

Table VIII

*Search tasks by experimental condition*

| Condition 1 goal: Minimal motivation - Minimal domain expertise. |
|---|
| Deze opdracht dient als oefening voor je vaardigheden m.b.t. het schrijven van een uiteenzetting.<br><br>Verzamel daartoe online informatie en schrijf een uiteenzetting van minimaal 500 woorden over het volgende onderwerp: 'Naties & Landsgrenzen in Centraal-Amerika in de vroege 19e eeuw.' |
| Schrijf voordat je begint kort op wat je van bovenstaande opdracht vindt: |

| Condition 2 goal: Maximal motivation - Minimal domain expertise. |
|---|
| Verzamel online informatie en schrijf een uiteenzetting van minimaal 500 woorden over het volgende onderwerp: 'De ontwikkeling van het geheugen van geboorte tot het 4e levensjaar.'<br><br>Binnen dit experiment zullen een aantal proefpersonen jouw uiteenzetting krijgen als startpunt voor hun eigen opdracht. |
| Schrijf voordat je begint kort op wat je van bovenstaande opdracht vindt: |

| Condition 3 goal: Maximal motivation - Maximal domain expertise. |
|---|
| Verzamel online informatie en schrijf een uiteenzetting van minimaal 500 woorden over het volgende onderwerp: 'Veranderingen in persoonlijkheidseigenschappen gedurende het leven.'<br><br>Binnen dit experiment zullen een aantal proefpersonen jouw uiteenzetting krijgen als startpunt voor hun eigen opdracht. |
| Schrijf voordat je begint kort op wat je van bovenstaande opdracht vindt: |

**APPENDIX D: Inter-Rater Agreement**

| Concepts and codes | | CP% * | PoA** | Cohen's κ |
|---|---|---|---|---|
| Relevancy evaluation | | 23 | 0,83 | 0,59 |
| | Evaluating possible relevancy | 11 | 0,87 | 0,42 |
| | Evaluating using relevant information | 6 | 0,96 | 0,75 |
| | Evaluating relevancy | 6 | 0,92 | 0,46 |
| Credibility evaluation | | 17 | 0,95 | 0,84 |
| | Evaluating source credibility | 6 | 0,96 | 0,71 |
| | Concerned about credibility | 5 | 0,99 | 0,90 |
| | Evaluating semantic credibility | 1 | 0,99 | 0,66 |
| | Evaluating surface credibility | 4 | 0,97 | 0,65 |
| | Using distrusted source | 1 | 0,99 | 0,66 |
| Undefined Evaluation | | 6 | 0,91 | 0,31 |
| | Evaluating possible usability | 4 | 0,96 | 0,39 |
| | Evaluating usability | 2 | 0,92 | -0,04 |
| Other activities | | 20 | 0,87 | 0,68 |
| | Copy/Pasting Paraphrasing | 7 | 0,95 | 0,75 |
| | Determining target information | 2 | 1,00 | 1,00 |
| | Sense-making | 2 | 0,96 | 0,55 |
| | Experiencing search difficulties | 5 | 0,97 | 0,82 |
| | Navigating | 0 | - | - |
| | Time management | 4 | 0,97 | 0,65 |
| | Saving information | 0 | - | - |
| | Translating | 0 | - | - |
| Information searching skills | | 34 | 0,87 | 0,73 |
| | Using multiple search engines | 0 | - | - |
| | Finding new search ideas | 0 | - | - |
| | Multilingual searching | 0 | - | - |
| | Entering prev. known sources | 3 | 0,99 | 0,79 |
| | Inferring content from abstract | 0 | - | - |
| | Search term selection | 10 | 0,96 | 0,83 |
| | Scanning information | 21 | 0,91 | 0,75 |

*Note:* CP% = Code Prevalence % in first coding session, and PoA = Percentage of Agreement. '-' indicates that PoA and κ were not calculated due to code prevalence being 0 in both coding sessions.

## APPENDIX E: Interview Quotations Regarding Search Engines

| Participant | Researcher Question | Participant Answer |
|---|---|---|
| 1 | Je begon eigenlijk direct door naar Scopus en Scholar te gaan. Waarom die specifiek? | Scholar is 1 van de sites waar ik eigenlijk altijd kijk omdat het een goede inspiratie is om artikelen te zoeken. Omdat je als mogelijkheid alles kan invullen en er zijn meestal goed dingen die je kan gebruiken. En Scopus is voor wetenschappelijke artikelen belangrijk, want het zijn meestal wetenschappelijke artikelen, Daar is de betrouwbaarheid van de informatie het best en bij Google Scholar. |
| 2 | Je gaf aan, vrijwel vanaf het begin, sprong je direct naar de Utwente site en daar naar web of knowledge en google scholar. | Ja zoals ik al zeg, als ik naar informatie kijk ga ik altijd naar UBTwente om te kijken welke site ik zal kiezen maar de keuze van de site op UBTwente was eigenlijk random. Google Scholar zoek ik eigenlijk het meest. |
| | Waarom dan juist die twee bronnen? Goede ervaringen? | Ja met Google Scholar heb ik goede ervaringen en die [Web of Knowledge] gebruik ik ook soms maar niet zo vaak. Maar dat was alleen omdat ik.....[zin niet afgemaakt] |
| 13 | Waarom begin je bij scholar of bij wetenschappelijke artikelen. Je had ook bij google kunnen beginnen. | Ergens in de opdracht [FOUT! Informed Consent] stond iets over betrouwbaarheid en op google dacht ik komen allerlei pagina's die niet zo betrouwbaar zijn dus daarom dacht ik: scholar. |
| | Als je nu bij google scholar iets zou zoeken waar je nog niet zo veel vanaf weet, hoe weet je dan dat die informatie klopt? Is dat omdat het google scholar is, dat je het dan eerder als betrouwbaar bestempeld? | Ja, meestal wetenschappelijke artikelen dus ja. |
| 14 | Waarom heb je voor scholar gekozen? | Ik denk het een omdat het een opdracht is voor academische vaardigheden, dat we ten eerste wetenschappelijke artikelen zullen moeten gebruiken. En ik heb dan voor scholar gekozen zodat ik wetenschappelijke artikelen kan verwerken voor de opgave. |
| | Dus je weet dat je daar wetenschappelijke artikelen gaat vinden die wellicht bruikbaar zijn? | Ja |

*Note:* Participants 1, 2 and 13 were in experimental condition goal 2: 'Maximal Motivation - Minimal Domain Expertise'. Participant 14 was in experimental condition goal 3: 'Maximal motivation - Maximal domain expertise'.

## APPENDIX F: Credibility Evaluation Quotes by Condition

Note that quotations in this section are taken out of context. Code assignment to quotations was performed by listening to the audio of the retrospective think-aloud exercise, watching their on screen actions, and reading their transcribed verbalizations of participants.

Table IX

*Condition goal: Minimal Motivation - Minimal Domain Expertise*

| Credibility Evaluation Codes | Participant Quotes |
|---|---|
| Using distrusted source | **Participant 7**<br>- Maar er was niets verder dus ben ik toch op Wikipedia gekomen.<br>**Participant 9**<br>- Nu probeer ik toch iets op wikipedia te vinden omdat ik geen andere idee had. |
| Concerned about credibility | **Participant 9**<br>- Dit lijkt geen serieuze site te zijn. |
| Source credibility evaluation | **Participant 7**<br>- En de eerste pagina's zijn eigenlijk alleen wikipedia dus daarom wil ik eigenlijk verder zoeken omdat ik niet op wikipedia wilde zoeken, omdat het misschien geen betrouwbare bron is<br>**Participant 9**<br>- Maar er zijn bijna wikipedia sites en die zijn niet zo serieus.<br>- Maar weer zijn er heel veel wikipedia sites en die wil ik niet.<br>- Nu probeer ik toch iets op wikipedia te vinden omdat ik geen andere idee had.<br>- Maar ik ben niet echt tevreden met de bron en vind ook niet echt de informatie die ik zoek.<br>- Ik kijk nu over toch een goede bron tussen de niet serieuze bronnen zijn.<br>**Participant 10**<br>- Toen heb ik voor bbp gekozen, omdat het een betrouwbare bron is uit duitsland zeg maar, maar dat was helemaal niet wat ik zocht. Dat was helemaal geen antwoord op mijn vraag, dus. |
| Semantic credibility evaluation | - |
| Surface credibility evaluation | - |

Table X

*Condition goal: Maximal Motivation - Minimal Domain Expertise*

| Credibility Evaluation Codes | Participant Quotes |
|---|---|
| Using distrusted source | **Participant 5**<br>- Ik denk dat er hier wel iets wordt verteld over de ontwikkeling en Wikipedia is niet echt een hele betrouwbare site of zo maar als je gewoon algemene informatie wil hebben dan staat er misschien toch nog iets nuttigs op. |
| Concerned about credibility | **Participant 1**<br>- Ik ben in het internet gegaan en heb vooral bij SCOPUS gekeken omdat er wetenschappelijke artikelen zijn en ook bij SCHOLAR, omdat het makkelijker is om daar ook andere informatie te verkrijgen [dan bij Scopus].<br>**Participant 3**<br>- Ik onthoud de pagina [ze opent een nieuw tab voor een nieuwe zoekopdracht] maar ik ga verder met een andere pagina. Want ik vertrouwde de pagina niet helemaal. Dus ik ben weer op zoek.<br>- Ik ben het nu aan het doorlezen, het lijkt me een vertrouwelijke [betrouwbare] bron en denk dat het dus wel relevant is voor mijn opdracht.<br>- Ik kijk naar de sites en welke vertrouwelijk [betrouwbaar] is, dat probeer ik uit te vinden. Ik ben nu naar de algemene informatie aan het kijken, de websitenaam en de informatie die er allemaal staat. Wat allemaal in het groot staat.<br>**Participant 5**<br>- Ik wist niet of de site heel betrouwbaar was bijv., maar omdat er gewoon wat algemene informatie op staat, ga ik gewoon even kijken wat er allemaal op staat. |
| Source credibility evaluation | **Participant 3**<br>- Omdat ik het toch niet zeker ben over wat voor website het is denk ik eraan om weer terug te gaan [naar Google] en andere informatie te zoeken.<br>**Participant 5**<br>- Ik denk dat er hier wel iets wordt verteld over de ontwikkeling en Wikipedia is niet echt een hele betrouwbare site of zo maar als je gewoon algemene informatie wil hebben dan staat er misschien toch nog iets nuttigs op<br>- Nou, gezondheidsnet, daar staat soms nieuws op en worden onderzoeken op gepubliceerd. Maar ik in hoeverre deze site betrouwbaar is dat weet ik ook niet echt. Dus het ligt een beetje aan wat er wordt verteld en wat er op staat.<br>- En dan wordt er straks ook nog iets verteld over een onderzoek van een hoogleraar, die psychologie geeft dacht ik, dus dat leek me dan wel betrouwbaar om mee te nemen in je eigen antwoord. Want dan krijg je dus ook antwoord op de vraag waarom je jouw eigen geboorte niet kan herinneren. |
| Semantic credibility evaluation | - |
| Surface credibility evaluation | **Participant 3**<br>- Ik ben op een site gekomen, het is van een boek denk ik. Het is betrouwbaar, ik zie bronnen en namen staan.<br>- Ook kijk ik naar de jaartallen van de bronnen, omdat het een vertrouwelijke bron is, het is literatuur, dus het zal wel vertrouwelijk zijn en daarom is het dan ook relevant voor mijn opdracht.<br>- Ook omdat er staat 'uit onderzoek concludeert' is voor mij een punt van dat kan ik wel noteren want dat is dan een feit, een conclusie van een onderzoek.<br>- De informatie leek namelijk op mijn eerder gevonden informatie uit een andere bron. Dus heb ik het er toch bij gezet. |

Table XI

*Condition goal: Maximal Motivation - Maximal Domain Expertise*

| Credibility Evaluation Codes | |
|---|---|
| Using distrusted source | **Participant 13**<br>- Hoewel wikipedia denk ik niet zo betrouwbaar is, maar alleen voor zo'n algemene uitspraak zal het wel goed zijn. |
| Concerned about credibility | **Participant 14**<br>- Daarna ben ik naar de browser gegaan en op google scholar wetenschappelijke artikelen te vinden over het onderwerp van persoonlijkheidsveranderingen.<br>- Geen wetenschappelijke artikelen, maar toch artikelen te vinden die ik kan gebruiken om de opdracht te doen.<br>**Participant 15**<br>- Dan kijk ik even naar de andere pagina, wat er van informatie staat. Ik vond deze website niet zo betrouwbaar op het eerste gezicht. |
| Source credibility evaluation | **Participant 11**<br>- Maar het was alleen een bijdrage van een vrouw waarvan ik niet wist of het wetenschappelijk was of niet.<br>- Ik heb nu voor een forumbijdrage gekozen, maar het was niet belangrijk omdat ik wist dat deze informatie niet wetenschappelijk zijn.<br>- En daarom ben ik op een pagina van het gezondheidsplein gegaan, omdat ik hier eerder wetenschappelijke informatie vind denk ik.<br>- Volgens mij is dit de beste pagina omdat dit het rijksinstituut van volksgezondheid is en die korte, maar goede informatie schreef. In de eerste zin stond eigenlijk al wat ik wilde weten. Of de persoonlijkheid veranderbaar is. Ik denk dat ik er de beste informaties uit heb gehaald.<br>**Participant 13**<br>- Hoewel wikipedia denk ik niet zo betrouwbaar is, maar alleen voor zo'n algemene uitspraak zal het wel goed zijn. |
| Semantic credibility evaluation | **Participant 11**<br>- Ik kies nu de website voor de big 5 omdat ik er al iets over gehoord heb heb en ik weet dat het wetenschappelijk is.<br>- En ik scan snel de informatie om te kijken of het ook juist is.<br>**Participant 13**<br>- Daaronder staat dan ook dat er op dit moment nog geen theorie over persoonlijkheidsontwikkeling is. Wat volgens mij niet echt klopt, en wat ook geen informatie oplevert voor deze opdracht. Dus is wikipedia in dit geval absoluut niet te gebruiken.<br>- Daar heb ik nog even snel doorgelezen wat er qua andere studies instaat en daar zie ik erikson staan. En dat is altijd goed want die heeft veel te maken met persoonlijkheidstrekken.<br>**Participant 15**<br>- Toen had ik twee bronnen waarvan ik dacht dat ze goed zijn. En keek ik eerst naar de eerste, het nationaal kompas van volksgezondheid en dat lijkt een betrouwbare bron omdat er goede informaties staan. |
| Surface credibility evaluation | - |

## APPENDIX G: Code Co-Occurence

In order to further determine the relationships between codes and to detect prevalent strategies that information seekers apply, all quotations were given a 'timestamp'. Timestamps followed from the final coding template and are largely mutually exclusive. Since every quotation was coded with a timestamp, it is possible to analyze what participants are wondering and doing during each phase of the information searching process by examining the codes that co-occurred with each timestamp. For each timestamp each code is listed if it occurred with the timestamp in at least 10% of the times the timestamp was coded in Table XII. The predominant co-occurring codes within each timestamp further support the conclusion that employing information searching skills to find relevant information is the primary concern of participants.

Table XII

*Timestamp - Code - Co-Occurrence >10%*

| Timestamp | Prevalence of timestamp | Co-occurence with code (>10%) | # of Code-Timestamp Co-occurences |
|---|---|---|---|
| Reading Assignment | 8 | Determining Target Information | **4** |
| | | Sense-making | 1 |
| URL Entry/Search Engine Selection | 13 | Entering Prev. Known Sources | **12** |
| | | Concerned About Credibility | 2 |
| | | Time Management | 2 |
| Search Entry | 58 | Search Term Selection | **54** |
| | | Multilingual Searching | 9 |
| | | Experiencing Search Difficulties | 8 |
| Examining Search Results | 56 | Evaluating Possible Relevancy | **31** |
| | | Evaluating Possible Usability | 9 |
| | | Scanning information | 8 |
| Examining Abstract/Introduction | 25 | Inferring content from abstract | **18** |
| | | Evaluating Possible Relevancy | **11** |
| | | Evaluating Possible Usability | 4 |
| | | Evaluating Using Relevant Information | 3 |
| Examining Source | 157 | Scanning information | **63** |
| | | Evaluating Possible Relevancy | **39** |
| | | Evaluating Relevancy | 30 |
| | | Evaluating Using Relevant Information | 30 |
| | | Copy/Pasting/Paraphrasing | 20 |
| Working on assignment | 49 | Evaluating Using Relevant Information | 5 |
| | | Copy/Pasting/Paraphrasing | 5 |

*Note:* Bold-faced numbers indicate a >20% co-occurrence rate between timestamp and code.