Impact of duration of the search on trust judgment of Wikipedia articles

Bachelor thesis by Tabea Hensel

Supervisors: Teun Lucassen, MSc. & Prof. Dr. Jan Maarten Schraagen

05-03-2012

Abstract

In the past years, Wikipedia has developed to one of the most visited websites. Due to its open editing character credibility evaluation can be a challenge for many users. Therefore different models have tried to assess credibility evaluation. Research on another construct, namely attitude certainty has shown that it is influenced by perceived evaluation duration of either familiar or unfamiliar topics. This study investigates whether the duration of the search also influences the trust judgment of Wikipedia articles with either familiar or unfamiliar topics. In the experiment, participants were asked to perform an information seeking task and to rate their trust in the articles. Analysis reveals that the duration of the search does not have impact on the trust judgment, whereas familiarity does.

Table of contents

1. Introduction

Due to the fact that Wikipedia is an open encyclopedia where anybody can edit anything, at the first glance, this encyclopedia seems less trustworthy than other online encyclopedias to many people. Evaluating the information before using it is therefore crucial. Although this seems logical it often happens that the user just takes the online information for granted without carefully thinking about it (Metzger, 2007).

Wikipedia was launched in 2001 and only ten years later, it scores the sixth rank of the most visited websites worldwide (ALEXA). Nowadays, nearly everybody who is working with the internet to gather information makes use of Wikipedia sooner or later. Although this collaborative encyclopedia scores one of the top ten ranks worldwide, its content is still disputable. Opponents of Wikipedia argue that through the collaborative authoring it is hard to evaluate the credibility of the articles. The fact that Wikipedia has a different quality control mechanism than other encyclopedias causes difficulties when evaluating the credibility. However, the study of Giles (2005) has revealed that there was no significant difference between the errors that the Wikipedia articles that were used for this study contained as opposed to the used articles from Encyclopaedia Britannica. In his study, it was shown that Wikipedia articles contained four errors on average, whereas those from Encyclopaedia Britannica contained three. Further criticism obtained by the opponents of the free online encyclopedia concern the missing visibility of the authorship. However, research has revealed that users often do not even scrutinize authorship, source and other important cues (Lim & Simon, 2011).

Anyhow, users certainly react on different cues from websites and make estimations. In the following, a definition of the concepts of trust and credibility is given. Furthermore, two different models of credibility assessment will be presented. These are the dual processing model of Metzger (2007) and the 3S-model of Lucassen and Schraagen (2011). Examples of factors that affect credibility evaluation are information skills (Lucassen & Schraagen, 2011) and motivation (Metzger, 2007). An additional factor that may affect the credibility evaluation is duration of search. Nobody has done research on a contextual factor as duration of the search so far. Therefore, it is the aim of this study to find out whether the duration of the search, influences the perceived credibility of either familiar, or unfamiliar topics of Wikipedia articles. This factor will be introduced and validated. After that, the method is introduced. Finally, the results are presented and discussed.

1.1 Credibility and trust

The words "credibility" and "trust" are often used interchangeably, as well as in common language as in research (Fogg & Tseng, 1999). On closer inspection, a difference becomes apparent. Believability can be used synonymously for credibility (Fogg & Tseng, 1999). The key elements of credibility are trustworthiness and expertise. The former can be described with three words: "well-intentioned, truthful and unbiased" (Fogg & Tseng, 1999, p. 40). Expertise, however, can be identified as "knowledgeable, experienced and competent" (Fogg & Tseng, 1999, p. 40). Besides these primary dimensions, other factors such as source attractiveness and dynamism have effect on credibility, too (Metzger, 2007). Users that try to assess credibility should keep five criteria in mind when evaluating a website: accuracy, authority, objectivity, currency and coverage (Alexander & Tate, 1999). Although these criteria are a helpful guidance for the credibility evaluation, research has shown that users hardly take these criteria into account when evaluating credibility. Surprisingly, research has revealed that design and the presentational elements often have a greater impact on perceived credibility of the user than the five criteria mentioned above (Metzger, 2007).

An important difference between credibility and trust is that trust implies a notion of dependability (Fogg & Tseng, 1999). Being dependent on the information also includes a certain level of risk. If there was no risk, trust would not be necessary at all (Kelton, Fleischmann & Wallace, 2008). Kelton et al. (2008) introduced a model for trust in digital information. They found the interpersonal ("I trust you") level of trust to be the proper one when talking about trust in information. In short, if you trust someone or something, you have confidence in it (Fogg & Tseng, 1999).

In this paper, credibility is seen as something that belongs to the message property, whereas trust can be seen as an attribute that belongs to the information user himself (Lucassen & Schraagen, 2011). In the following, the term credibility evaluation refers to the process of assessing credibility. In the experiment of this study, participants are asked to rate their trust in the information and confidence in their judgment among other things.

1.2 Models approaching trust assessment

Different models and theories have been introduced to describe the process of credibility evaluation and important factors that influence the credibility evaluation. Two of

these models are the dual processing model of credibility assessment (Metzger, 2007) and the 3S-model recently introduced by Lucassen and Schraagen (2011). The dual-processing model of website credibility assessment is anchored in the dual processing theory by Chaiken (1980). Motivation and the ability to evaluate are central in Metzger's model which is split up into three phases: the exposure phase, the evaluation phase and the judgment phase. In the first phase, the question emerges whether the user is motivated and has the ability to evaluate. If the user is motivated *and* has the ability to evaluate, in the evaluation phase, it will lead the user to a systematic/central evaluation. If the user is motivated but the ability to evaluate is missing, there are two possible outcomes. This leads either to no evaluation at all, or to a heuristic/peripheral evaluation. In case of a heuristic or systematic evaluation, the user makes a credibility judgment in the judgment phase. This model accentuates the process of information seeking, which fits to the information seeking task of this study.

Lucassen and Schraagen (2011) introduced a more detailed model for trust in information, named the 3S-model. The name derives from the three features of information. These are semantic features, surface features and source features. Accuracy and completeness are examples for semantic features. Features that focus on the surface are for example length and pictures. Source features refer to the experience the user has made with the source. According to the 3S-model, trust judgment is based on user characteristic and information characteristics. Expertise and source experience are both part of user characteristics. There are two types of expertise, namely domain expertise and information skills. The former refers to the level of expertise of the user. The latter is related to the skills the user has to evaluate credibility. Trust judgment can also depend on the source experience. The user then is referring to previous experience with the source of information. While basing the judgment upon the source experience active evaluation is not demanded. However, domain expertise and information skills require an active credibility evaluation

1.3 Additional factor: duration of the search

Lucassen and Schraagen (2011) have assessed important user characteristics that influence the trust judgment of online information. Nevertheless, every judgment is also influenced by contextual factors such as mood or motivation (Metzger, 2007). In the study of Metzger, Flanagin and Medders (2010), participants have reported to stop with their search once they have found information that served as confirmation for their own knowledge.

Therefore the question arises whether, besides the user and information characteristics, a contextual factor, for instance, the duration of the search for information, has impact on the credibility evaluation as well.

Tormala, Clarkson and Henderson (2011) have done research on the effect of perceived evaluation duration (PED) on attitude certainty. PED can be defined "as the subjective speed with which one generates an evaluation of an object" (Tormala et al. 2011, p.422-423). Research on PED and attitude certainty has revealed two findings. First, when evaluating familiar objects, fast evaluation (short duration of evaluation) leads to more certainty. Second, when evaluating unfamiliar objects or forming an opinion, slow evaluation (long duration of evaluation) leads to more certainty. The former can be described as more heuristic (peripheral), while the latter seems more systematic (central).

Although attitude certainty is not the same as a trust judgment the core might be similar. Taking a closer look at attitude certainty and trust, it is striking that both have one thing in common: confidence. Attitude certainty can be defined as "a subjective sense of conviction, confidence, clarity or correctness one has about one's attitude" (Tormala et al., 2011, p. 422). Trust, too, contains the notion of confidence (Fogg & Tseng, 1999). Therefore, it seems possible that evaluation duration may influence trust in information, too. *Perceived* evaluation duration already implies that the subjective evaluation duration was measured. However, the present study assesses the objective and thus actual duration of the search for information. This study aims to investigate whether it is possible to translate the findings of the research on PED and attitude certainty into trust in information.

1.4 Hypotheses

Tormala et al. (2011) have revealed that the perceived evaluation duration and the familiarity with a topic have impact on attitude certainty. Applying these findings to trust in information may reveal the same results.
This leads to the following hypotheses:

As research on PED (Tormala et. al, 2011) has shown, when evaluating a familiar topic, a quick search positively influences attitude certainty.
H1: Quick search positively influences the trust judgment of familiar topics.
It seems that the familiarity positively influences the trust judgment. If people are familiar with the topic, they judge it more credible (Self, 1996; Chesney, 2006).

When evaluating an unfamiliar topic, quick search negatively influences attitude certainty. People may feel the need to evaluate an unknown object more carefully. Therefore the second hypothesis will be tested.

H2: Quick search negatively influences the trust judgment of unfamiliar topics.

2. Method

2.1 Participants

A total of 40 college students took part in this experiment. All participants were behavioral science students. The students could enroll online to participate in this experiment. In return, the students received course credits. Some participants were invited verbally or via email. The average age was 21 ($SD = 2.48$). 14 male students (35%) and 26 female students (65%) participated. Of those 40 students, 26 of the participants were German, 13 of them were Dutch and one person was Greek. It was assumed that all participants were equally proficient in the English and Dutch language as their education at the university is given in both English and Dutch.

2.2 Stimuli

The participants of this experiment received eight different Wikipedia articles that originated from the English Wikipedia. Four articles dealt with familiar topics, four with unfamiliar topics. The articles were of B-class quality, as defined by the Wikipedia Editorial team. This means that they are nearly complete and do not contain any remarkable faulty information, but to reach the "good article status" style and content needs to be improved. B-class quality was the average quality of the articles. Instead of the real online Wikipedia articles, screenshots from the articles from Wikipedia were taken to disable the search function and links. A commercial at the top of the page and the silverlock[1] at the top right corner of the website were removed. The commercial was removed because it appears only temporarily, has no benefits and might distract. It was decided to remove the silverlock because participants might not know its meaning and could become confused. Furthermore, nothing else was adjusted but kept original. It was not possible to use the search function during the task. Participants were assumed not to be aware of modifications.

---

[1] http://en.wikipedia.org/wiki/Wikipedia:Protection_policy

2.3 Task

The participants had to perform an information seeking task on a computer in a laboratory. Lime Survey, an online survey tool, was used for the questionnaire. The participants were presented with eight Wikipedia articles. For each article, a question concerning the content of the article was given. One example of the questions that were asked about the article concerning Cannabis was: How much THC is at least required to have a perceptible psychoactive effect from cannabis? Participants were asked to find the correct answer in the article. The correct answer to the question could always be found in the article. After each article, they were asked to rate their trust and confidence in the article and how familiar they were with the topic. A brief motivation for their trust judgment was expected, as well.

2.4 Procedure

Before the actual task began, the participants were asked to sign an informed consent. After that, they received instructions for the actual task. According to the instructions, demographic information and their habits concerning their Wikipedia use needed to be filled in by the participants. Once the participant continued, it was not possible to return to the previous page. After all, people could optionally leave remarks.

2.5 Design

The design of this experiment is a 2 (familiarity: familiar/unfamiliar) x 2 (location of information: top/bottom) within-subjects design. The order of the articles was counterbalanced.

2.6 Independent variables

2.6.1 Familiarity

Looking for differences between the trust ratings of familiar and unfamiliar topics, familiarity was manipulated. A short pretest with four people was conducted to distinguish familiar from unfamiliar topics. The familiar topics were Cannabis, Greenday (band), the Little Mermaid (film) and Operant Conditioning. Unfamiliar topics were Digestion, the Elasmosaurus, Sao Tomé and Principe and Simon Stevin.

Familiar topics were selected with regard to the age of the participants and their study background. The expected average age was 20, so topics like famous bands and films were chosen. Cannabis is a hot issue in the Netherlands which is why it was assumed that people would be familiar with this topic, too. Due to their educational background, the article about operant conditioning was selected. Unfamiliar topics were chosen concerning the specialty of the topics. All topics that were categorized as unfamiliar were assumed not really to be of broad interest and were very specific.

2.6.2 Duration of search

The participants received questions that needed to be answered by using information that could be found in the article itself. The correct answer could be found either at the top or at the bottom of the page. So, the location of the information was not manipulated directly, but the questions that were asked were either related to the first third of the article or the last third. Placing the sought information either at the top or at the bottom of the Wikipedia article will urge the participant to put less or more time into reading the article.

2.7 Dependent variables

2.7.1 Trust ratings

After each article, participants had to rate their trust in the article on a 7-point Likert scale ranging from 1 (very low trust) to 7 (very high trust).

2.7.2 Motivations for trust judgments

Besides the trust ratings on a 7-point Likert scale, participants were asked to provide a brief motivation for their trust judgment. These motivations were coded afterwards. The coding scheme was developed based on the 3S-Model of Lucassen and Schraagen (2011). Three categories were made: semantic features, surface features and source features. An additional category, other motivations, was added in case the motivation did not fit in any of the three categories. In table 1 the subcategories are presented. After the analysis of the motivations, Cohen's Kappa was calculated. 20% of the motivations were double coded by two persons. Cohen's Kappa had a value .75, which indicates a substantial agreement.

Table 1 Subcategories of the features

Semantic Features
Accuracy
Completeness
Scope
Neutrality
Structure
Other semantic features

Surface Features
References
Internal Links
External Links
Pictures
Length
Writing Style
Appearance
Topic
Other surface features

Source Features

Other motivations

2.7.3 Confidence ratings

Additionally to the trust ratings, the level of confidence in the trust rating was measured on a 7-point Likert scale ranging from 1 (extremely low confidence) to 7 (extremely high confidence). Due to the fact that confidence resembles attitude certainty, this variable was added.

2.7.4 Familiarity ratings

Participants were also asked to assess their familiarity with the topic of the article on a 7- point Likert scale ranging from 1 (extremely low familiarity) to 7 (extremely high familiarity). An ascending value indicates higher familiarity with the topic.

2.7.5 Time

Time that was spent on reading each article was measured in seconds. It was assessed to check how much time was needed to find the correct answers to the questions. The time

measurement began as soon as the participants viewed the articles and ended as soon as they had filled in an answer to the question that was presented on top of the screen.

2.8 Data analysis

A Wilcoxon signed-rank test was conducted to test whether the manipulations of familiarity was successful. A paired sample t-test was conducted for the manipulation check of the duration of the search. To test the hypotheses, a repeated-measures ANOVA was performed. For the analysis of the motivations, a chi-square test was conducted. Furthermore, a two-proportion z-test was conducted to compare the proportions of the motivations.

3. Results

3.1 Manipulation checks: familiarity & time

Familiarity and duration of the search were manipulated. Analysis of the familiarity ratings showed that there was a significant difference between the ratings of familiar and unfamiliar articles, $Z = -5.24$, $p = 0.000$. The participants rated the familiar topics higher ($M = 3.76$, $SD = 0.89$) than the unfamiliar topics ($M = 1.68$, $SD = 0.49$).

Also the manipulation of time, respectively the location of the information, was successful. Participants needed more time to answer questions where information was at the bottom of the page ($M = 407.99$, $SD = 27.96$), than at the top ($M = 245.04$, $SD = 15.89$), $t(39) = -6.55$, $p = 0.000$.

3.2 Trust ratings

No main effect for duration of the search was found for the trust ratings, $F(1, 39) = 0.67$, $p = 0.417$. No interaction between familiarity and duration of the search was found either, $F(1, 39) = 0.00$, $p = 1.00$. A main effect for familiarity was found on the trust ratings, $F(1, 39) = 19.03$, $p = 0.000$. The participants' trust ratings were higher on the familiar topics ($M = 5.56$, $SD = 0.10$), than on the unfamiliar topics ($M = 5.03$, $SD = 0.12$). This was shown by post-hoc inspection. None of these findings support the two hypotheses. Therefore, the hypotheses were rejected.

Further analysis showed that there were differences between the trust ratings of the eight articles. Post hoc inspection showed that the article about Cannabis was trusted most ($M = 5.78$, $SD = 1.00$), whereas the article about the Elasmosaurus was trusted least ($M = 4.37$, $SD = 1.21$).

3.3 Confidence ratings

Analysis of the confidence ratings revealed the same effects as the trust judgment analysis. The main effect of location was non-significant, $F(1, 39) = 0.85$, $p = 0.363$. Again, a main effect for familiarity was found, $F(1, 39) = 11.404$, $p = 0.002$. Confidence in articles with a familiar topic was higher ($M = 4.96$, $SD = 0.17$) than in articles with an unfamiliar

topic ($M = 4.51$, $SD = 0.17$). No interaction between the duration of the search and the familiarity with the topics was found, $F(1, 39) = 0.91$, $p = 0.347$.

3.4 Motivations for the trust ratings

All participants were expected to fill in a motivation. That resulted in a total of 320 comments. One participant filled in random letters four times. This might have been either because he did not know what to fill in, or to finish the experiment more quickly. The rest of the comments were categorized in four categories. These were: Semantic features, surface features, source features or other motivations. The category "semantic features" included statements about the accuracy, completeness, neutrality, scope, structure or other semantic features. One example of accuracy is *"I'm familiar with such topics. I haven't seen any mistakes, so to me it seems credible"*[2].

Motivations that alluded to the "surface features" were references, internal and external links, pictures, length, writing style, general appearance of the article, topic or other surface features. Statements like *"A lot of literature, external links, references to the book"*[3] scored both on references and external links. "Source features" referred to the experience people have made with Wikipedia and their attitude towards it. An example of such statement is *"Image Wikipedia (so many contributors)"*.[4] Statements that did not really fit to any of the other categories were categorized in "other motivations". *"I don't know"*[5] is one example of other motivations.

Chi-square tests were conducted in order to find out whether there were differences in the distribution of the used features of the four categories. Due to the fact that there was only one significant main effect, the analysis of the motivations focuses on the familiar and unfamiliar topics of the articles.

A Chi-square test showed that there was no significant difference between the motivations in the four categories for familiar and unfamiliar topics $\chi^2 (3, N = 464) = 4.37$, $p = 0.225$.

---

[2] Translated from Dutch "Ben wel bekend met dit soort themas. Geen fouten gezien, dus wel betrouwbaar naar mijn oordeel"

[3] Translated from Dutch "Veel literatuur, externe links, referrence to the book"

[4] Translated from Dutch " Imago Wikipedia (so many contributors) "

[5] Translated from Dutch "Weet ik niet"

Further analysis revealed that there were marginally significant differences between the 17 subcategories χ^2 $(14, N = 400) = 29.75$, $p = 0.08$. There was a significant difference between familiar and unfamiliar topics for three features. These were accuracy, pictures and length. Accuracy was mentioned significantly more often with familiar topics (5.4%) than with unfamiliar topics (1.7%).

"Pictures" were mentioned significantly more often with unfamiliar topics (5.8%). The feature "length" was used by 4.6% as motivation for unfamiliar topics, whereas only 1.3% of the motivations dealt with this features for familiar topics. The percentages from the trust ratings were illustrated in table 2.

The rest of the used features did not differ significantly for familiar or unfamiliar topics. The feature that was referred to the most for both, familiar and unfamiliar topics was "references". For familiar topics, 29.5% of the motivations were about references. For unfamiliar topics, 22.1% of the motivations referred to references.

Tabel 2 Percentages and totals of the motivations

	Familiar	Unfamiliar
Semantic features		
Accuracy	5.4% (n=12)	1.7% (n=4)
Completeness	4.5% (n=10)	2.1% (n=5)
Scope	.4% (n=1)	.8% (n=2)
Neutrality	2.7% (n=6)	.8% (n=2)
Structure	5.8% (n=13)	7.5% (n=18)
Other semantic features	10.7% (n=24)	11.3% (n=27)
Surface features		
References	29.5% (n=66)	22.1% (n=53)
Internal Links	.9% (n=2)	2.1% (n=5)
External Links	1.3% (n=3)	3.8% (n=9)
Pictures	.9% (n=2)	5.8% (n=14)
Length	1.3% (n=3)	4.6% (n=11)
Writing Style	4.9% (n=11)	5.4% (n=13)
Appearance	6.7% (n=15)	8.8% (n=21)
Topic	8.9% (n=20)	6.7% (n=16)
Other surface features	2.2% (n=5)	2.9% (n=7)
Source features	2.7% (n=6)	.8% (n=2)
Other motivations	11.2% (n=25)	12.9% (n=31)
Total	100% (n=224)	100% (n=240)

3.7 Other findings

The trust judgments and the level of confidence in their judgments showed a positive correlation, $r = 0.91$, $p = 0.002$, for the eight different articles. This was calculated by the means of the eight articles. On the individual level, the correlation between trust and confidence was $r = 0.47$, $p = 0.002$. For the individual level, the mean of the eight articles for each participant for both, trust and confidence, was calculated.

Comparing the correlations between trust and confidence in the familiar and unfamiliar topics, analysis showed that the correlation is higher for familiar topics $r = 0.38$, $p = 0.000$, than for unfamiliar topics $r = 0.29$, $p = 0.000$.

The correlation of all trust and confidence judgments was $r = 0.366$ ($r = 0.366$, $p = 0.000$).

4. Discussion

In this paper, it was hypothesized that both, the familiarity with the topic and the duration of the search have impact on the credibility evaluation. It was expected that quick search would (1) positively influence the trust judgment of familiar topics and (2) negatively influence the trust judgment of unfamiliar topics. However, the results of this study suggest that the duration of the search has no effect on the trust judgment. No main effect for the duration of the search was found. An interaction effect for familiarity and the duration of the search wasn't found, either. Due to these findings, all hypotheses needed to be rejected. Nevertheless, a main effect for familiarity was found for the trust judgments. The trust judgments of familiar and unfamiliar topics differed from each other. Trust was higher in articles with a familiar topic than it was in articles with unfamiliar topics. Furthermore, a positive relationship between trust scores and confidence scores was found, both on the individual level as well as in the eight articles.

The main effect for familiarity supports the findings of Self (1996) and Chesney (2006). They both found that people make more positive trust judgments when rating familiar topics. A possible explanation for this effect may be that when people are familiar with a topic, they can base their judgment upon their own knowledge. Presented facts can be verified more easily when knowledge about a topic is present. If one is not familiar with a topic, one mainly has to focus on surface or source characteristics and has to make a credibility evaluation on the basis of these features (Lucassen & Schraagen, 2011). This is confirmed by the finding that accuracy was mentioned more frequently with familiar topics than with unfamiliar topics, assuming that accuracy can only be determined when one's own knowledge can be brought to bear.

The familiarity effect may also be caused by the length of the articles. In this experiment, articles were not supposed to be equal except for the quality. Characteristics such as number of pictures or references of the articles were not taken into account to keep the experiment as realistic as possible. However, familiar articles turned out to be twice as long as unfamiliar articles. Participants may have made the inference that due to the length of the text of unfamiliar articles, the authors might have put more effort in writing the text (Palfrey & Gasser, 2008). This was confirmed by the finding that the feature "length" was mentioned more frequently with unfamiliar articles than with familiar articles. Hence, the trust judgment may have been influenced by the length of the article, rather than by the familiarity with the topic.

Participants' trust judgment did not seem to be influenced by the duration of the search. This could be due to several reasons. One possible explanation for the absent effect of the duration of the search is that the first impression of an article might interfere with a more elaborate credibility evaluation. This means that within a few seconds, people might have already formed their opinion about an article, regardless of the familiarity with the topic or the time they spent reading. Even if the user is spending another few minutes on reading the article, he or she might only try to find confirmation for this impression. Metzger et al. (2010) has found this behavior as participants from their study reported to look for confirmation of their own knowledge.

Another explanation may be that in the study of Tormala et al. (2011) the perceived evaluation duration was measured differently. The subjects from Tormala et al. (2011) received false feedback that led them think that evaluating something has either taken them more or less time than other participants. In the present study, the time starting at first sight of the article until the leaving of the article after answering the question was measured. Two different constructs were measured. Whereas in this study the actual time was measured, participants in the other study were simply made to believe that they needed more or less time. The results suggest that there is a difference between subjective and objective evaluation duration. It might be that participants were not aware of duration of their evaluation until they received the feedback. Therefore, the effect might be caused by heuristics rather than duration.

In this study, a familiarity effect was also found for the confidence ratings. Again, confidence was higher for familiar topics than unfamiliar topics of the articles. So, participants were more confident of their trust judgment for articles with a familiar topic. Further evidence for this relation was found through a correlation analysis. A positive correlation between trust in information and confidence in the trust judgment was found. Unsurprisingly, since confidence is one facet of trust. As Fogg & Tseng expressed in 1999, trust indicates a certain amount of confidence. People might dare to make a positive trust judgment more easily when they have at least some level of expertise and are thereby more confident of their judgment. These findings are totally in line with the findings of the trust ratings from this study and provide additional support for the construct of trust and credibility evaluation.

4.1 Limitations and future research

The articles could have been chosen better in relation to the equality of articles. Familiar articles were nearly twice as long as the unfamiliar ones on average. Although articles were intentionally chosen to differ in many ways but the quality, the proportion between familiar and unfamiliar articles could have been more equal. It is not impossible that the main effect for familiarity was evoked by the length of the articles. Further research should focus on the equality of the features to find out what caused the main effect for familiarity.

Furthermore, an experiment about Wikipedia might make people more suspicious than they usually are. Asking people explicitly to make trust judgments also affects the judgment. As soon as people know that they are asked to give reasons for their trust judgments, they may try to answer socially desirable.

It would also be advisable to use another group of participants for further research than people who participate in psychological experiments regularly. These people probably may assume to know what the experiment is about and consequently do not behave naturally.

4.2 Conclusion

This study has not provided any proof for the hypotheses. In this study, the duration of the search had no impact on the trust judgment. Further research needs to be done to validate this finding. However, trust judgment was influenced by familiarity. Trust was higher in articles with a familiar topic, rather than with unfamiliar topics. Furthermore, a strong relationship between trust and confidence was found. Additional support for the 3S-model from Lucassen and Schraagen (2011) was provided, too.

5. References

Alexa Top 500 Global Sites, (2012). Retrieved 3 March, 2012, from
http://www.alexa.com/topsites

Alexander, J. E. & Tate, M. A. (1999). *Web Wisdom; How to Evaluate and Create
Information Quality on the Web*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1st
edition.

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(5),
752-766.

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*,
11(11).

Fogg, B.J. & Tseng, S. (1999). Credibility and Computing Technology. *Communications of
the ACM*. 42(5), 39-44.

Giles, J. (2005). Internet encyclopedias go head to head. *Nature*, 438(7070), 900-901.

Kelton. K., Fleischmann, K.R. & Wallace, W.A. (2008). Trust in digital information. *Journal
of the American Society for Information Science and Technology, 59* (3), 363-374.

Lim, S & Simon, C. (2011). Credibility judgment and verification behavior of college
students concerning Wikipedia, *First Monday*, 16 (4).

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(7), 1232-1242.

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(13), 2078-2091.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibilty Evaluation Online. *Journal of Communication* , 413-139.

Palfrey, J. & Gasser, U. (2008). *Born Digital: Understanding the First Generation of Digital Natives* (1st ed.). New York: Basic Books.

Self, C.S. (1996). Credibility. In M. Salwen & D. Stacks (Eds.), An integrated approach to communication theory and research. Mahwah,NJ: Erlbaum.

Tormala, Z. L., Clarkson, J. J., & Henderson, M. D. (2011). Does fast or slow evaluation foster greater certainty? *Personality and Social Psychology Bulletin*, 37(3), 422-434.