Physiological Measurement of Trust-Related Behavior in Trust-Neutral and Trust-Critical Situations

Karin Leichtenstern, Nikolaus Bee, Elisabeth André, Ulrich Berkmüller, and Johannes Wagner

Human Centered Multimedia, Universitätsstr. 6a, D-86159 Augsburg
{Leichtenstern,Bee,Andre,Wagner}@hcm-lab.de

Abstract. In this paper we present results of a user study that we conducted with 21 subjects to investigate whether initial user trust is accompanied by unconscious bodily responses which enable more objective measurements than user reports. In particular, we recorded the user's eye gaze and heart rate to evaluate whether users respond differently when interacting with a web page that is supposed to build initial trust as opposed to a web page that lacks this capability. Our results indicate that there are significantly different response patterns to trust-critical and trust-neutral situations during the interaction with a web page depending on whether the web page has helped users form initial trust or not. Knowledge of trust-related behavior can help to manage user trust at the runtime of the system since different usage phases can continuously be interpreted in order to detect situations which need to be considered to re-cover user trust.

1 Introduction and Related Work

Trust is a major factor to ensure that a computing system will find acceptance among users. For human to human and similarly for human to computer relationships, we learn to appreciate other people or computing systems and thus, in process of time, we acquire knowledge about their actions and behavior. This is a basis for trust and a mechanism to reduce social complexity [4]. If a human does not have previous knowledge and experience in another human or computing system, a social dilemma emerges since humans are often forced to interact with other unknown humans or things. In this context, we talk about *initial trust* [7] if a user is able to form trust based on her first impression towards an unknown interaction partner. For instance, when interacting with web pages in the Internet, users often do not have previous knowledge about the web pages and their vendors. If the web page does not succeed in building up initial user trust, the user will probably avoid using it and its services. Research has focused on these issues and other problems related to trust in several disciplines, such as

sociology, psychology, economy and computer science (e.g. [8]). Typically, these different disciplines have several perspectives on trust and consequently definitions. One common agreement is that trust is multi-dimensional consisting of subjective trust triggers, such as the user's willingness of vulnerability, benevolence, reliability, competence, honesty, and openness. These subjective triggers can lead to different levels of attention and engagement as perceived states of user trust in a computing system [3]. There is evidence that trust increases the users willingness to acquire new knowledge (e.g. [6]) since they overcome the feelings of insecurity and risk [8]. We therefore hypothesize that a web page which succeeds in establishing initial user trust is also more likely to motivate users to study its content which potentially can be measured by increased levels of attention and engagement. Various studies indicate that attention and user engagement can be measured using physiological data (e.g. [10]).

In this paper we investigate whether the perceived user behavior of initial trust can be measured by making use of objective measuring methods. Applying observation methods instead of inquiry methods can solve two major issues: (1) Not the user's subjective trust opinion is measured by inquiry methods but instead their actual trust-related behavior by observation methods which potentially provide more realistic and genuine user data. (2) Also, the user's trust behavior is not assessed after the usage but continuously while interacting with the system. By having access to body sensors (e.g. wearable sensors, such as as a pulse meter embedded in clothes) and external sensors (e.g. an eye tracker), the user's attention and engagement might be assessed and thereby indirectly the level of trust. Then, if necessary, the level of user trust can be re-covered by self-adaptively optimizing the system, such as by providing more system transparency or user control. In the following we present a first step towards investigating these trust-related behaviour. We present a user experiment where 21 users were interacting with two web pages in a trust-neutral and a trust-critical phase. One of these web pages was expected to support the formation of initial trust while the other was not. While interacting with the two web pages, physiological data of the user's heart rate (HR) were captured as well as their eye gaze behavior. Later on, the data was analyzed by means of SPSS and discussed how trust-related behavior by means of attention and engagement can be measured objectively.

2 Implementing Initial Trust

In order to objectively investigate initial trust in web pages, we decided to implement two versions of a web page which pretended to provide free software downloads, such as games or office tools. One web page called *WorldofApps* was expected to support initial trust since it followed approved trust guidelines whereas the other web page called *LoadIt* did not support these trust guidelines and thus potentially lacked initial trust. The applied trust guidelines are mainly based on recommendations of Nielsen et al. [9] and Fogg et al. [2]. Among other guidelines, we addressed guidelines in terms of the web page's credibility

Fig. 1. WorldofApps that potentially forms initial user trust (left) and LoadIt that potentially lacks initial trust (right)

by providing contact information and an imprint. Additionally, we considered guidelines with respect to usability, such as an intuitive and easy to use navigation. Since initial or immediate trust is strongly linked to an appropriate and professional appearance [5], we also took into account a calm and serious layout and design for the web page *WorldofApps* (see Figure 1 left) whereas the web page *LoadIt* provided several blinking images and texts (see Figure 1 right). The reliability of the user interface was also covered by dead links for *LoadIt*.

3 The Experiment

The main objective of the experiment was to investigate whether trust is accompanied by typical heart rate and eye gaze patterns. Therefore, we used our two introduced web pages which potentially form initial trust (*WorldofApps*) or not (*LoadIt*) and conducted an experiment.

As independent variable we defined *the applied user interface* with two levels: user interface with initial trust and user interface with a lack of initial trust. The dependent variable was *the initial user trust* in the user interface which was investigated subjectively and objectively. We decided to also subjectively measure trust by a questionnaire in order to validate whether *WorldofApps* was more likely perceived as trustworthy than *LoadIt* and thus whether the two web pages could form initial user trust or not. Our questionnaire based on common trust dimensions which are considered as factors of trust (e.g. [1]). The subjects had to rate their perceived level of the trust dimensions (correctness, security, reliability, credibility, basic usability and appearance) for the corresponding web page on a five point scale from strongly disagree to strongly agree. Additionally, we also asked for a rating of the web page's overall trust level. Thus, we also added a statement in terms of the overall user trust. For objectively measuring the users' behavior we applied sensors for measuring heart rate and an eye tracker to record the users' eye gaze behavior. As group design, we applied the within subjects design and thus each user had to participate in both levels of the independent variable. To prevent positioning effects we counterbalanced the order of the levels.

In the experiment we asked 21 participants (two female and 19 male) with a strong professional background in computer science to complete different tasks with our two web pages and *Amazon*. They were aged between 21 and 27 ($M = 24.8$, $SD = 1.97$). Before we started the experiment, we asked the subjects for a permission to capture physiological and audio-visual data while they were completing their tasks. Then we equipped them with our sensors, re-calibrated the eye tracker and started the capturing. The experiment always started by first filling-in a questionnaire with questions addressing demographics and then by browsing *Amazon* as a period of familiarization. We also used *Amazon* for the users to relax and get used to browsing the Internet for our setting. The participants had to browse *Amazon* for new products for approximately two minutes. Afterwards, the subjects had to fill-in our second questionnaire regarding statements in terms of the subject's perceived trust dimensions and overall user trust for *Amazon*. We also asked for trust ratings of *Amazon* since we expected these ratings as a typical baseline for a trustworthy web page. Followed by the period of familiarization, eleven subjects started by using *WorldofApps* and later on used *LoadIt* while the other ten subjects used *LoadIt* first. For both web pages, the subjects first were asked to freely browse on the web page for about two minutes. We consider this browsing phase as a trust-neutral phase since it did not contain serious trust-critical situations. After the browsing phase, the subjects had to download two predefined software products from the respective web page which took on average 55.4 seconds for *WorldofApps* and 75.4 seconds for *LoadIt*. In contrast to the browsing phase, the download phase contained serious trust-critical situations since the computer might be harmed by the download of a corrupted software product. That is the reason why we call the download phase a trust-critical phase. After the browsing and download phase the participants had to fill-in our questionnaire for the corresponding web page.

4 Subjective Validation of Initial Trust

The results of the questionnaire validate that our participants perceived *WorldofApps* as trustworthier than *LoadIt* (see Fig. 2). A one-way repeated measured ANOVA test showed significant results in all categories: correctness ($F(2, 40) = 73.56$, $p < 0.001$), security ($F(1.4, 27.5) = 94.14$, $p < 0.001$ with sphericity corrections by Greenhouse-Geisser), reliability ($F(2, 40) = 92.5$, $p < 0.001$), credibility ($F(2, 40) = 125.59$, $p < 0.001$), basic usability ($F(2, 40) = 55.13$, $p < 0.001$) and appearance ($F(2, 40) = 75.80$, $p < 0.001$). For each trust dimension the Bonferroni post-hoc tests revealed significant differences between the web pages of *Amazon* and *LoadIt* as well as between the web pages *WorldofApps* and *LoadIt* with $p < 0.001$. For the trust dimensions about security ($p < 0.001$) and credibility ($p < 0.05$) we also had significant differences between *Amazon* and the web page *WorldofApps*.

In terms of overall user trust, we also found highly significant differences between the three web pages ($F(2, 40) = 144.30$, $p < 0.001$). The Bonferroni post-hoc test revealed significant differences between *Amazon* and *WorldofApps* ($p = 0.001$) and between these two web pages and *LoadIt* (each with $p < 0.001$).

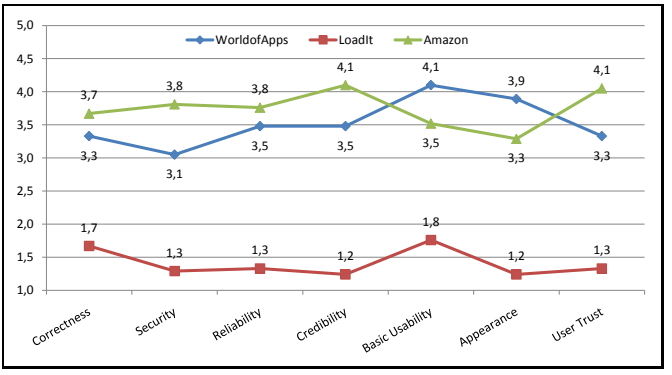

Fig. 2. Results of the Questionnaire (Trust Dimensions and User Trust)

As expected we observed that *WorldofApps* and *Amazon* achieved similar high values of the trust dimensions and the overall user trust which confirms our assumption that the web page *WorldofApps* helps forming initial user trust. In contrast to these two web pages, *LoadIt* failed in all categories which seems to indicate a low level of initial user trust. The bad results of *LoadIt* also appeared when analyzing the feedback of the subjects. Several subjects mentioned that they were reluctant to download software from the bad web page since they were afraid about also downloading a virus or a trojan. One subject even refused to download software from *LoadIt*.

5 Results of the Objective Data

The captured physiological data was split into three sections each with a duration of 30 seconds. The first section (*Browse Begin*) began 15 seconds after the browsing task started and the second section (*Browse End*) ended 15 seconds before this task was finished. The third section (*Download*) began 15 seconds after the download task started which was directly following the browsing task. We decided to use the offset of 15 seconds for each section to reduce interfering factors which are caused by the transition from one phase to another. The trust-neutral phase (browsing task) is covered by the first and second section and the trust-critical phase (download task) is covered by the third section. For the analysis of the objective data we applied two-sided pairwise t-tests. One of our 21 subjects was excluded for the analysis of the objective data since some problems occurred with the sensors during the capturing.

The eye tracker (SMI iView X RED) recorded the users' eye gaze with 50 Hz. We computed the average fixation times from the raw gaze data within the single sections (see Fig. 3). We could not find significant differences between the fixation time of the trustworthy web page in comparison to the web page with little initial trust. However, the fixation time computed in milliseconds during the section *Browse Begin* of the trustworthier web page ($M = 353, SD = 118$) was higher (1) compared to the web page with little initial trust ($M = 294$,

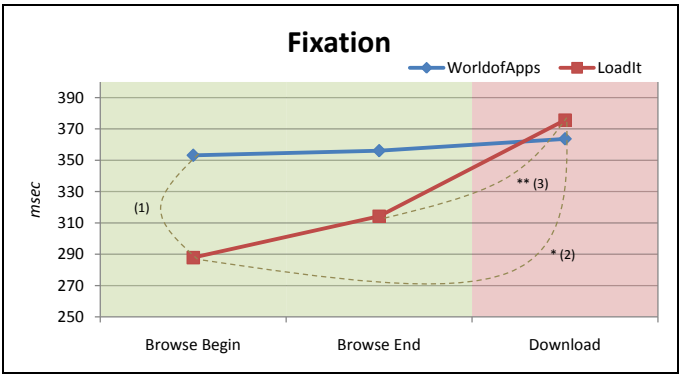

Fig. 3. Fixation time in milliseconds during the three sections. *Green* area indicates the trust-neutral phase and the *red* area the trust-critical phase. ($^{*}p < 0.05$, $^{**}p < 0.01$).

$SD = 96)$, $t(18) = 1.8, p = 0.085, r = 0.40$ with a medium effect size. Within the three sections of *LoadIt*, we found a significant increase (2) between *Browse Begin* ($M = 287, SD = 100$) and *Download* ($M = 376, SD = 110$), $t(18) = -2.6, p < 0.05, r = 0.53$ with a large effect. And further we also found a significant increase (3) between *Browse End* ($M = 310, SD = 77$) and *Download* ($M = 376, SD = 110$), $t(18) = -3.3, p < 0.01, r = 0.61$ also with a large effect size. Overall, the users devoted more continuous attention to the web page that succeeded in building initial user trust than to the web page that failed to do so. Only in trust-critical moments, namely when users had to download software the level of attention increased for the little trustworthy web page and achieved the level of attention of the trustworthy page.

The heart rate was computed by measuring the length between the interbeat intervals and averaged within the three sections (see Fig. 4). Comparing the

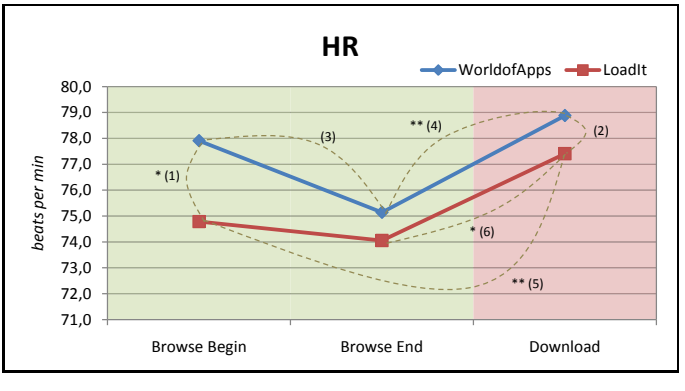

Fig. 4. Heart rate in beats per minute during the three sections. *Green* area indicates the trust-neutral phase and the *red* area the trust-critical phase. ($^{*}p < 0.05$, $^{**}p < 0.01$).

heart beats per minutes of *WorldofApps* ($M = 77.9, SD = 10.8$) with *LoadIt* ($M = 74.8, SD = 10.6$) during *Browse Begin* reveals a significant decrease (1) with a large effect size ($t(19) = 2.5, p < 0.05, r = 0.50$). A decrease (2) between *WorlofApps* ($M = 78.9, SD = 9.7$) and *LoadIt* ($M = 77.4, SD = 9.4$) also appeared during *Download* with a medium effect size ($t(19) = 2.0, p = 0.060, r = 0.42$). Within the three sections of the trustworthy web page, the heart rate was higher (3) at *Browse Begin* ($M = 77.9, SD = 10.8$) compared to *Browse End* ($M = 75.1, SD = 9.0$) with a medium effect size ($t(19) = 1.9, p = 0.077, r = 0.39$). After *Browse End* there was a significant increase (4) towards *Download* ($M = 78.9, SD = 9.7$) with a large effect size ($t(19) = -3.5, p < 0.01, r = 0.63$). Within the three sections of the web page without initial trust, the heart rate was significantly lower (5) at *Browse Begin* ($M = 74.8, SD = 10.6$) compared to *Download* ($M = 77.4, SD = 9.4$) with a large effect size ($t(19) = -3.5, p < 0.01, r = 0.63$). And there is also a significant increase (6) between *Browse End* ($M = 74.1, SD = 9.1$) and *Download* ($M = 77.4, SD = 9.4$) with a medium effect size ($t(19) = -2.4, p < 0.05, r = 0.48$). Based on the results it seems that higher values of the heart rate go along with higher engagement with a task completion. This higher engagement, in turn, seems to be an indicator of a user interface's success to build initial user trust, particularly, since the pattern of higher heart rates appeared independent on whether the users were interacting in a trust-neutral (browsing) or trust-critical (downloading) phase. Furthermore, we found a significant pattern if users changed from a trust-neutral to a trust-critical phase which appeared similarly for a web page that succeed in building initial user trust and also for the web page that failed to do so. Consequently, the heart rate seems to be also an indicator if users change in situations that require more user trust. A further insight is that a trustworthy web page seems to have a relax phase when browsing (trust-neutral phase) since the heart rate decreased until the end of the browsing phase. In contrast to that, the phenomena of a relax phase was not observed for the web page with a lack of initial trust.

6 Conclusion

Based on the results of our experiment we found significantly measured patterns of the user's eye gaze fixation time and heart rate when interacting with a user interface that was subjectively rated as trustworthy or untrustworthy. Additionally, we found significant patterns if users complete trust-neutral tasks (browsing) or trust-critical tasks (downloading) for the trustworthy and untrustworthy rated systems. The results of our user study indicate that users do not take user interfaces with a lack of initial trust seriously which turned out to be reflected by reduced attention and engagement to complete a task. Once the user has changed from a trust-neutral (browsing) to a trust-critical phase (downloading), the user's attention and engagement was significantly increased when using a systems that lacks of initial trust. In this situations the users seem to be forced to change their behavior towards an increased level of engagement and attention for not being harmed by the system. As far as we know we are the first who did investigations of the user's actual trust-related behavior by means

of physiological data. Our work provides a first step towards the development of methods that continuously measure user trust based on behavioral data. Such methods bear the advantage over traditional trust measurements that they enable a more objective assessment of user trust. They are not based on subjective self reports which have to be provided by users either after the interaction or cause an interruption of the experience. During the usage time, our methods provide important knowledge about the users and their behavior in order to automatically manage user trust by means of self-adaptations, such as by presenting more system transparency and user control. In future work we would like to validate our results by means of a setting that has several changes between trust-critical and trust-neutral situations.

Acknowledgement

This research is sponsored by *OC-Trust* (FOR 1085) of the German research foundation (DFG).

References

1. Chen, S., Li, J.: An empirical research on consumer trust in e-commerce. In: IEEC 2009: Proceedings of the 2009 International Symposium on Information Engineering and Electronic Commerce, pp. 56–61. IEEE Computer Society, Los Alamitos (2009)
2. Fogg, B.J., Marshall, J., Laraki, O., et al.: What makes web sites credible?: a report on a large quantitative study. In: CHI 2001: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 61–68. ACM, New York (2001)
3. Hughes, L.W., Avey, J.B., Norman, S.M.: A study of supportive climate, trust, engagement and organizational commitment. Journal of Business & Leadership: Research, Practice and Teaching 4(2), 51–59 (2008)
4. Luhmann, N.: Vertrauen; ein Mechanismus der Reduktion sozialer Komplexitat. F. Enke, Stuttgart (1968)
5. Lumsden, J.: Triggering trust: to what extent does the question influence the answer when evaluating the perceived importance of trust triggers? In: BCS HCI 2009: Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction, pp. 214–223. British Computer Society (2009)
6. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. The Academy of Management Review 20(3), 709–734 (1995)
7. McKnight, D., Cummings, L., Chervany, N.: Initial trust formation in new organizational relationships. The Academy of Management Review 23(3), 473–490 (1998)
8. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: An integrative typology. Info. Sys. Research 13(3), 334–359 (2002)
9. Nielsen, J., Snyder, C., Molich, R., Farrell, S.: E-Commerce User Experience. Nielsen Norman Gr. (2001)
10. Pecchinenda, A., Smith, C.A.: The affective significance of skin conductance activity during a difficult problem-solving task. Cognition & Emotion 10(5), 481–504 (1996)