TIME: A Method of Detecting the Dynamic Variances of Trust

Laurian Vega, Yeong-Tay Sun, D. Scott McCrickard, Steve Harrison

Center for Human-Computer Interaction
2202 Kraft Drive
Virginia Tech, Blacksburg, Virginia, USA

{laurian, ytsun, mccricks, srh}@vt.edu

ABSTRACT

Given that interactions are dynamic, we propose that trust is also a dynamic, unfolding, and a deeply contextual phenomenon that must be evaluated as such. The central argument of this paper is that there is a need to measure trust iteratively and in situ. This measurement of trust can provide a deeper insight into the construct of trust and the design elements that influence it. In this paper we present a review of trust evaluation methods. We then propose our method, the TIME Method, using repeated measures of trust across multiple pages of a website to tie design elements to increases or decreases in user trust. We then evaluate user trust with the TIME Method to demonstrate the degree of trust variability. Last, we discuss future methods for evaluating trust.

Categories and Subject Descriptors

H.5.M. [**Information Interfaces and Presentation**]: Miscellaneous.

General Terms

Measurement, Experimentation, Human Factors

Keywords

Trust, credibility, empirical evaluation, measurement, TIME Method, dynamic evaluation.

1. INTRODUCTION

Considerations of trust in the design of websites are often centered on design elements or heuristics that will increase or decrease a user's trust. Design with this focus can be critical in areas like eCommerce [32] and eHealth [37] where user decisions are found to rely heavily on constructs like trust. The most accessible methods of evaluating user trust have come to rely heavily on post interaction measurements (i.e., using Likert-scale questions). Something these methods do not address is the important part that micro interactions - the click-by-click user interactions - can play in changing a user's trust level and subsequently the user's decisions.

Given that interactions are dynamic, we propose that trust is also a dynamic, unfolding, and a deeply contextual phenomenon that must be evaluated as such. Post-interaction questionnaires alone cannot address the evolution of user trust over time. The central argument of this paper is that there is a need to measure trust iteratively and in situ. This measurement of trust can provide a deeper insight into the construct of trust and the design elements that influence it. We propose our method in this paper, called the TIME (**T**rust **I**ncremental **M**easurement **E**valuation) Method. The TIME Method places an emphasis on repetition, taking repeated measures of trust across the pages of a website with many users. Subsequent analysis forms the basis for identifying relationships between specific design elements and increases or decreases in user trust. In this paper we provide two contributions. The first is a review of trust evaluation methods. The second is a discussion of an initial implementation of the TIME Method to demonstrate the degree of trust variability. Last, we discuss future methods for evaluating trust.

WICOW'10, April 27, 2010, Raleigh, North Carolina, USA.

2. BACKGROUND

Trust is an area of research with a dense background stemming from fields as diverse as psychology [7], decision making [33], medicine [24], computer science [1], and others. Trust is a multidimensional construct that conveys a multitude of meanings and uses. Similar to credibility, trust is a user-held perception based on a history of interaction with an environment. However, the distinctions between the two terms are not always clear. In one conceptualization, trust is part of the construct of credibility, as notably explained in the work of Hovland in 1953 [16]. As Fogg and Tseng explain, credibility is the appraisal of believability made by assessing the trustworthiness and expertise of a source. Trustworthiness is composed of the sub-constructs of well-intentioned, truthful, and unbiased [13].

Alternatively, trust can be argued as an assessment of a larger socio-technical system of which the assessed source is only a part. In this conceptualization, credibility is part of the assessment of trust. In determining the credibility of a health webpage, for example, the assessment of trust will depend on several factors within the larger socio-technical system. A doctor's referral to a particular webpage or the availability of personal health records embedded within the webpage are both contributing factors to page credibility and to overall trust in the system. For this work, we have focused on assessing the user's trust, with an understanding of how credibility assessments may additionally be affected.

This work explores how users assess trust in a website. We have focused on one particular form of a website: information repositories composed of user-generated content. These websites were specifically selected because they have intrinsic issues with source credibility and authority that can affect a user's assessment of trust. For example, in 'Can you ever trust a wiki' the authors explain how the reliability of online information repositories (e.g.

Wikipedia) leads to distrust due to the mutability of the webpage content [23]. The credibility of online information is further called into question when key visual cues in the interface are missing. Lumsden and MacKay's study showed that participants use visual signals, like contact information and VeriSign certificates, as social indicators of trust [26]. These visual signals are frequently missing in websites composed primarily of user-generated content.

There are models of how trust in online information is formed. One example model from Kelton and Fleischmann extends a general iterative model of trust to incorporate trust in online information [21]. They explain that by performing a search for online information, the user is entrusting the Internet and creating a situation of dependence. Their model shows how an information source is evaluated based on aspects of trustworthiness like accuracy, objectivity, validity, and stability. Other factors affecting trust include disposition, the relevance of information, and recommendations from authoritarian entities. While the topic of our work emphasizes the importance of micro-interaction effects, models of trust like this one are critical to developing a well-rounded understanding of trust.

Prior works have explored relationships between user interface elements and user assessments of trust (see [15, 31, 34] for examples). A goal of future work is to identify design elements with pronounced impacts on user trust. However, the work presented in this paper comprises the first step in characterizing the construct of trust, be it relatively stable or highly volatile.

Finally, other methods of analysis have examined the contextual in situ assessment of related constructs, although these methods have not been applied to trust. For example, the method of context-aware experience sampling can require a participant to carry a beeper device. When the beeper goes off the participant responds to some predetermined survey questions [6, 19]. While this method is similar, user interactions are surveyed across a much longer period of time (i.e., 4 times a day) [36] in comparison with the relatively brief periods in which micro interactions occur.

3. Method of Evaluation and Measurements of Trust

Trust as a social science grapples with a host of interrelated concepts, giving rise to a parade of domain-specific models; a few of these are discussed in the sections above. Trust has been quantified through the use of sophisticated calculus-based equations, where specific types of trust are represented using one or more related factors or antecedents [20]. Other models choose to prioritize theoretical concerns in addition to numerical validity, mindful of the conceptualization of trust as a complex, multifaceted construct [8]. Trust in online communities can be inferred through user observations: for instance, one's willingness to cooperate and work with peers, including those with no prior familiarization [18]. In software and electronic environments, trust can be broken down into system-specific considerations such as environmental integrity and information disclosure, which are then subjected to individual compliance assessments by a domain expert [1].

With so many different faces of trust, the significance of a trust measurement tends to be highly dependent on the questions being explored. Not surprisingly, it is difficult to apply a conceptual model without obligatory tinkering to mitigate practical limitations. This involves striking a delicate balance between models that remain conceptually valid, yet readily lend themselves to experimental endeavors. An example comes from the work of Kelton et al., described above [21]. They build their model of trust, and how to evaluate it off of earlier works by Mayer et al. [29] on trust in interpersonal relationships. Specifically, they narrow the concepts of trustworthiness between the models from competence, positive intentions, ethics, and predictability to accuracy, objectivity, validity, and stability.

Quantifying measurements of trust, in particular, is troublesome due to the large number of potential factors. Consider the challenges involved in appropriating calculus-based trust models. Equations of trust are highly problematic in that they often incorporate several different trust types. One is frequently difficult to distinguish from another outside of the conceptual realm.

Another strategy for evaluating trust is the creation of constructs in which participants interact. For example, in the work of Hassanein and Head [15], three webpages were created to display different levels of the construct 'social presence.' Their hypothesis was that increasing social presence would increase user trust. The webpage with low social presence would have a picture of items for sale with only a functional description. The webpage with high social presence had pictures with people present along with snippets of emotional user evaluation of the product. While manipulating specific aspects of an interface might reveal useful information about trust in websites, it detracts from the ecological validity of the findings; these are not real webpages. Additionally, these studies function by simply directing users to a webpage. This subtracts all of the rich contextual information surrounding the navigation and evaluation of that webpage. Last, these studies remove a level of dependence, which can be critical for trust. Users have no stake in whether or not their decisions are correct.

One of the most popular methods of trust measurement is the use of quantitative questionnaires. Combining the flexibility of multiple topics with Likert-style standardization [2, 3, 43], the main drawback of this technique is its reliance on user recollections of trust after the fact [36]. This can be troublesome,

	Method	Example Papers
Qualitative	Open-ended Questionnaires	Fisher et al. 2008
	Diaries	Sillence et al. 2007c
	Observations	Rosenbaum et al. 2008
	Open-ended Questions & Interviews	Sillence et al. 2007a; Tio et al. 2007
	Website & Email Logs	Sillence, Briggs et al. 2007c; Iacono and Weisband 1997
	In-person Surveys	Semere et al. 2003; Greenhalgh et al. 2008
	Group Interviews & Discussion	Sillence et al. 2004; Marshall and Williams 2006
Quantitative	Surveys	Hassanein and Head 2004; Bakker et al. 2006
	Interviews	Khoo, Bolt et al. 2008
	Online Questionnaires	Huntington et al. 2004; Walther et al. 2004
	Phone Surveys	Menon et al. 2002; Rosenvinge et al. 2003

Table 1. Example papers using different methods of evaluation to detect user trust in websites.

as studies have shown individual recollections of trust to be notoriously fallible, even when one's perception of trust is based from first-hand experience [9]. These types of studies can be general surveys that are mailed or emailed out to mass groups of people (see [17, 41] for examples), or they can be surveys given during a lab-based study. On the far end of this spectrum, mass surveys are highly de-coupled from the activities that form trusting behaviors. These surveys can be used to gather general impressions of trust but lack the ability to explore some specific factors (e.g., interface elements). At the other end of the spectrum, using surveys before and after user task completion still de-couples behaviors and trust assessments. However, it does allow for the experimenter to take more focused measurements with less recollection bias. Similar quantitative trust gathering methods include post-interaction participant interviews [22], and phone interviews [30, 34] where answers correspond to Likert-scale responses .

Alternatively, qualitative methods can be used to analyze a user's trust. Such methods can provide ideas about the factors that affect trust. Examples include the use of open-ended qualitative questionnaires [10], examining interpersonal communication such as email [18], and the use of diaries to document feelings of trust towards websites [39]. Other qualitative-like methods, such as the examination of web-logs [39], and in-person surveys [14, 35], can provide numerical counts that add to arguments in relation to trust and use.

Another qualitative approach to measuring trust is by analyzing observable behavior (e.g., [33]). A large proportion of research evaluating trust involves the use of a game similar to prisoner's dilemma. Prisoner's dilemma forces people to trust each other so that both teams receive optimal payment. The game works by having participants work together (or sometimes with a confederate) with each player having a set amount of goods. With both participants contributing goods to a job, they can both receive payout. If one team member betrays another, however, that individual has the opportunity to make more money. Researchers can thereby measure trust according to whether or not the players betray each other. This research makes the assumption that participants are rational and that the game elicits enough initial trust to establish a trusting relationship not worth betraying. An extended analysis of the prisoner's dilemma game can be found in (Riegelsberger et al. 2003).

There has been other qualitative work examining the effects of communication methods on team collaboration and trust (i.e. video chat vs. telephone). How the different teams perform provides an analysis of how different methods of communication impact trust [4]. While our study focuses on how an individual trusts a website, this method can be valuable when considering web communication methods.

Perhaps more compelling are interpersonal interviews [38, 40], group discussions [28], and direct observations of users performing system tasks. Such tasks are typically designed to elicit trust-related feedback on any number of system areas or features. These methods, while far more adaptable than inflexible equations, require a time commitment dedicated to study design, task preparation and analysis. Further, their highly qualitative nature is far more likely to introduce analytical biases and researcher-dependent interpretations. For example, the work of Sillence et al. [37] elicited fifteen users to come to an internet café and participate in a one-hour session gathering medical information. After a break, the participants then participated in fifty-minutes of group discussion about their experience. This allowed the researchers to determine what impact, if any, that the design and content of the website had on a user's trust.

The dichotomy of trust evaluation strategies supports the argument for a mixed method. As explained by Riegelsberger and Vasalou [32], "… trust can not solely be investigated with interviews or questionnaires, because individuals want to appear as rational decision-makers." While their argument centers on decision-based evaluation methods, we propose a method of evaluation where in situ repeated measurements of trust can be paired with a user's online information retrieval actions, thereby promoting decision process transparency.

4. DO MICRO INTERACTIONS AFFECT TRUST, AND IF SO, HOW TO MEASURE IT?

How often does a user's trust change and when is that change significant? Prior work proposes that trust can change when navigating between single-topic web pages [5], across a period of time such as a set of days [42], or when experiencing a shift in the interaction density [25]. Despite these transitions, however, measurements of trust are often delayed following the completion of a series of tasks. If trust is truly dynamic and not a simple net aggregate, then waiting until the end of an interaction or task to take a measurement may unwittingly mask the antecedents or affecting factors responsible for variation in a user's level of trust.

To these ends, we propose a method of measuring trust multiple times across multiple web pages within a single website. Part of this method shares similarities with the popular think-aloud protocol. While having the user think aloud or provide regular measurements may disrupt or possibly even change the user's activities, these methods are tools that can provide valuable insight while the user is engaged.

To fulfill our goals, a number of issues were considered. The first was to decide how often to perform the repeated measurement. Querying the user about her trust level required an interruption of the user during her task. Our goal was to time these interruptions frequently enough to accumulate sufficiently detailed measurements of the micro interactions without excessively distracting the user. To achieve this, a buffer of no less than thirty seconds was maintained between the previous user-triggered questionnaire and the next. This time interval was implemented to limit participant frustrations by establishing a sense of predictability to the interruptions while keeping the time between measurements reasonably discreet.

The second consideration was how to pair a user's trust with her actions. Data collection requires participants to be queried about their levels of trust despite trust being a highly subjective and implicit construct. For this reason, we paired the explicit trust evaluations with the more implicit beliefs informed by the user's behavior. Two strategies were employed for this purpose. First, we associated user interactions with individual clicks detected within the interface. Clicking is performed for numerous reasons but it generally implies that the user wishes to explore some other available webpage or web resource. We logged information about interface components being clicked on along with applicable trust measurements. We also paired a user's decisions by capturing video of the interaction as displayed on the screen. This allowed us to review the footage to match individual trust evaluations with specific system interactions of interest.

Choosing when and where to display the questionnaire form was a crucial decision in designing the evaluation method. (See Figure 1 for a depiction of the form in use.) It was our design goal to take measurements during periods of interaction, but not to overly interrupt the participant. Initially we considered polling the participants every thirty seconds, but periods where users were reading or writing down notes could not be accommodated. For this reason, we focused on users performing clicks in order to minimize disruptions to the current micro-task. The form would appear only if the current click was more than thirty seconds after the last time the form was displayed.

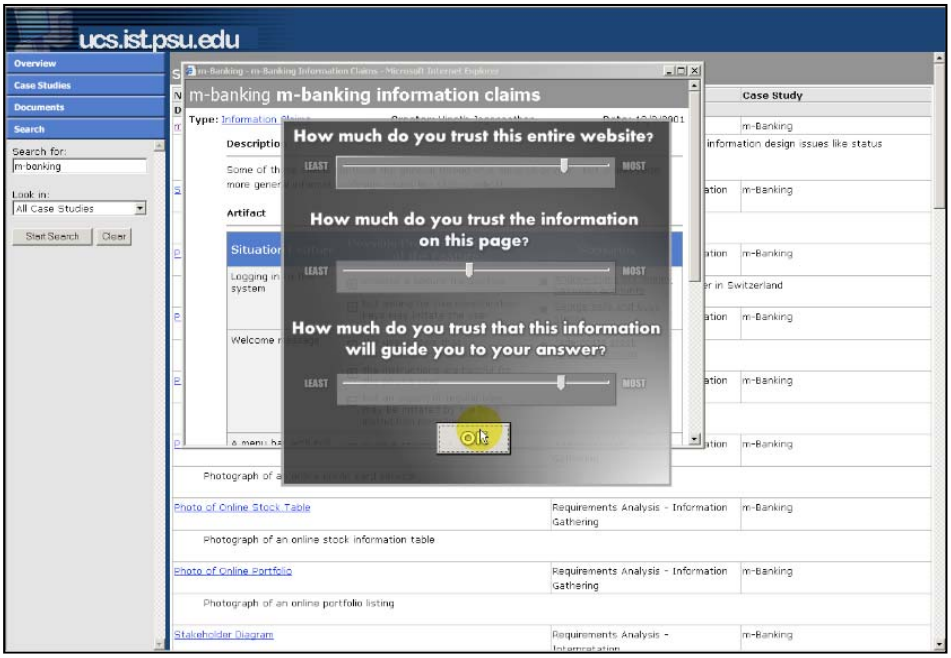

Figure 1. This picture shows a screenshot of one participant interacting with the Usability Case Study Library after the trust evaluation form has appeared

Additionally, we addressed specific methodology issues involved with taking measurements. We employed a continuous visual scale with values from zero to a thousand (unlabeled to the user). While the details in the range were not visible to the user, it allowed for a high level of precision. This large range made even small changes measureable as well as highly visible. Initially appearing as blank for each new query, clicking within a scale would reveal a visual trust marker to be freely manipulated by the user. Taking measurements in this way decreased the chance of identical consecutive measurements since information about the previous value was not available. It also reduced errors due to anchoring. For instance, if the prior measurement was visible, a user might feel that her current trust level did not warrant a change to the previous value – that the prior value was 'good enough.'

The last consideration we addressed was determining what the user was trusting. Building off the work of Flanagin and Metzger [11] and that of Chopra and Wallace [11], the results of trust evaluations can be different depending on what is being assessed and its hierarchy within the web-system. For this reason we wanted to see if trust evaluations and their variability would change depending on what was being trusted. To evaluate the TIME Method we used information repositories with user-generated content. We asked questions about trust for three levels:

Trust in the collective system. This scope of user trust is interesting in that it can approximate the severity of individual user experiences, good or bad, on the web-system as a whole. At this scope, it is easier to identify "deal-breaking" interaction occurrences that leave lasting impressions on users. This scope is also useful for comparing initial levels of user trust across multiple websites and how they evolve over time.

Trust in a small system subset. Smaller scopes of trust confined to well-defined contexts tend to evoke highly directed user feedback based on immediately available resources. Will functional or technical gaffes be magnified or lessened depending on specific web-system contexts? Frequency of occurrences? Or something else?

Trust in the small web-system subset as an indicator of progress. This question asks the user whether he or she believes the system is capable of supporting his or her activities, and is closely related to the design of information. If users are misled or frustrated too often, perceptions of system reliability may decrease. Alternately, indicating the attainability of information where none exists may produce similarly undesirable effects.

Participants were asked three questions. The first question asked how much the participant trusted the website as a whole. The second question asked how much the participant trusted the particular webpage she was viewing. The last question asked how much the participant trusted that the information she was viewing would guide her to answer her task. These three hierarchies demonstrate how trust may be different yet interconnected across different levels of a web-system.

An assumption of our evaluation method is that user behavior is goal directed. Users will seek out and navigate an information repository when they are looking for information. This assumption draws upon the subordinate relationship between trustor and trustee, sustained by the following two conditions. First, the trustor must have a need to fulfill. Information is one such need. Second, the trustee must be able to fulfill those needs. In this case, an information repository presumably contains the answers to our goal directed questions [21].

5. METHODOLOGY

5.1 Assessment of the TIME Method

Since our goal was to assess the TIME Method, we chose a proven experimental methodology: task guided interaction in online web-systems. One limitation of using task driven interaction is that the data from this study can only provide insight into similar activities. However, many activities that people perform with information repositories involve looking up information quickly - highly similar to the tasks completed by

study participants. For this study we limited the number of information repositories to two. This decision was made because of the limited time we had with participants, as it would be difficult to sustain participant engagement in studies lasting much longer than an hour. We also wanted to assess whether or not our method would be valuable.

We used a convenient sample of available students from the area of Human-Computer Interaction. For this reason the two information repositories with user-generated content selected were seen as representative knowledge bases relevant to the participants' education. The first information repository used by participants was the Usability Case Study Library (UCS). The UCS is a focused collection of case studies with associated design artifacts. The UCS is a real tool that is used to help educate HCI students about design in action. The content of the website is created by designers of all levels, including novices. The second information repository used by participants was Wikipedia. Wikipedia is an expansive, popular, and collaborative, user-editable online encyclopedia.

5.2 Demographics

Twenty people participated. Twelve were male. Participants were university students who had taken a Human-Computer Interaction class (9) or were graduate students in HCI. All participants accessed the Internet daily.

5.3 Procedure

Upon arrival participants completed a demographics questionnaire. The participants were told that they would interact independently with two different websites containing user-generated content. For each website, twenty minutes would be allotted to solving four website specific tasks, with a short break in between. During their tasks a form would periodically appear on-screen to solicit current levels of trust. The form could be moved around the screen so that the focus of evaluation was not obscured. The participant could not continue with the task before answering each set of trust evaluation questions.

The tasks took up the entire twenty minutes. It was emphasized that completing all of the tasks was not as important as completing a task correctly. An example task was: "Who was the first female presidential candidate?"

Undergraduate participants received extra course credit for participation. The study was piloted with two participants, resulting in minimal changes to the task wording. The data from the pilot was not included in the results.

6. RESULTS & DISCUSSION

The results have been organized into two sets: results that pertain to taking repeated trust measurements with the TIME Method, and results pertaining to the participants assessment of trust in the two information repositories.

6.1 Trust Measurements

Using the TIME Method to evaluate trust in both information repositories, the questionnaire was presented on average to the participants every 39 seconds. No statistical significance between the UCS and Wikipedia was found. The average participant was asked to fill out the form approximately 59 times. In total, there were 1,018 responses to the three trust questions. On average each participant answered the trust questions for the UCS twenty-six times, and for Wikipedia twenty-five times.

The average amount of time spent answering the trust questions was 10 seconds (variance = 22.4, standard deviation = 4.7). On the participants' first encounters with the trust question form they spent an average of 25 seconds answering the questions; this was during the UCS task portion of the experiment. While working with Wikipedia, participants spent 14 seconds answering the trust questions for the first time. The minimum amount of time answering the trust questions was 4 seconds per form answer, with a maximum of 47 seconds. To verify that participants were not simply clicking randomly to quickly bypass the form, two researchers checked for overt inconsistencies in the videos and graphs of responses. Participants spent on average approximately 4 minutes 27 seconds (22%) answering the form for the UCS during the 20-minute period and 3 minutes 38 seconds (18%) answering the form for Wikipedia.

All participants used the entire twenty minutes to complete the four tasks. Only one participant completed all tasks. There were no gender or education effects on the number of times answering the questionnaire. Three participants completed the tasks out of order. All participants had previously used Wikipedia; five participants had previously used the UCS. No statistical differences were found between those who had previously used the UCS and those who had not.

6.2 Trust Measurements with TIME

Gender and education effects were observed in trust values. Females were more trusting across all trust questions and for individual trust questions ($p<0.0146$). Undergraduates were also more trusting for all questions and individual trust questions ($p<.0001$).

In a comparison of the two information repositories, Wikipedia was statistically more trusted than the UCS (UCS Trust = 495, Wikipedia Trust = 617, p= 1.1E-35). This finding held across individual questions. Scores ranged from 0 to 1000.

If trust in a web-system is fairly stable and less susceptible to influence from small interactions, then variance should be minimal. However, the variance for the three trust questions was found to be significant. The average standard deviation of the participants' responses using the UCS were as follows: website = 102.0, webpage = 129.8, present information = 205.8; for Wikipedia: website = 78.6, webpage = 113.3, present information = 200.8. This result indicates that, on average, a participant's

Time (Level)	Difference	Trust Mean
0-5min	A	564.07
5-10min	B	489.03
10-15min	B	463.75
15-20min	B	443.83

Time (Level)	- Time (Level)	Difference	Lower CL	Upper CL
0-5min	10-15min	120.23	59.71	180.76
0-5min	15-20min	100.32	38.06	162.49
0-5min	5-10min	75.04	12.50	137.59
5-10min	10-15min	45.19	-17.46	107.84
5-10min	15-20min	35.27	-39.06	89.61
15-20min	10-15min	19.91	-42.46	82.30

Table 2. Statistical results from Tukey-Kramer HSD test for detection of time differences. A time difference was detected between first 5 minutes as denoted by A v. B.

response would increase or decrease approximately 10 percent when responding to how much they trusted the UCS as a whole website, and similarly for the rest of the values.

The second way we evaluated whether the TIME Method of evaluation was valuable was by detecting a temporal effect. The twenty-minute session for each information repository was divided into five-minute segments as well as by task. Trust measurements for one participant at different times can be seen in Figure 2. A time effect for both the UCS and Wikipedia was found using an all pairs Tukey-Kramer HSD test. For both the UCS and Wikipedia, the first five minutes of trust evaluations were significantly different from the remaining 15 minutes; trust was higher. The analysis can be seen in Table 2. This difference held when analyzing the data by website. The trust means in five-minute segments for the UCS were 1: 519, 2: 422, 3: 407, 4: 400; for Wikipedia, 1: 610, 2: 535, 3: 533, 4: 486.

7. DISCUSSION

Numerous measurements were taken during participant use of both websites. The first set of measurements pertains to the effectiveness of the method being used. Participants were asked to respond to the form at intermittent intervals exceeding 30 seconds. This satisfied our goal of focusing on periods of user interaction while avoiding the interruption of users engaged in processing web information. Despite these limitations, a sufficient number of trust measurements were attained.

In regards to how long participants spent responding to the form, two things may explain the higher than average times: first, participants wanted to become familiar with questions they had not encountered before. Second, participants wanted to create a mental benchmark for answering the question repeatedly. After answering the first set of trust questions, the time spent answering subsequent occurrences quickly stabilized for each participant. This second interpretation explains why participants spent about half the time responding to the form the first time when it was presented within the second information repository's tasks.

The participants spent between 22% and 18% of their time answering the questions on the form instead of completing their tasks. While this is a significant portion of the time, the participants were told that the primary goal of their participation was to assess their trust in the web system and that they were not being graded on whether or not they completed the assigned tasks. While this result reduces the ecological validity of the assessments somewhat (i.e. users are generally not subjected to constant polling while searching for online information) it is a side effect of taking repeated measurements. Alternative choices are to have a lower number of questions on the form or to have the form appear at less frequent intervals.

Gender and education did not affect the number of measurements taken but it did affect the trust value. People who are less educated and female have been found to trust more than people who are more highly educated and male. This finding is corroborated in other similar trust studies (see [12, 27]). This finding indicates that gender and education level may not affect the frequency of interaction with an information repository but it can affect a participant's subsequent assessment of trustworthiness.

On average and across all trust questions, participants trusted Wikipedia more than the UCS. This finding has two possible explanations. The first is that Wikipedia has more user interface features that enrich the interaction experience. It was observed that participants using Wikipedia experienced success in using the search feature and enjoyed intuitive browsing. Contrary to that experience, participants using the UCS had difficulty using the browser and were confused by the left-handed navigation menu. Usability problems have been documented to impact a user's trust [10]. The second explanation is that even though the effects of familiarity were not enough to affect the results of this study, it did have an impact. Work from Zahedi and Song have found that temporal and familiarity have an effect in relation to trust and health websites [42].

Seeing the variation of trust responses across consecutive pollings is the largest indicator of how micro-interactions can affect a user's trust assessment. This finding suggests that trust within even a twenty-minute period of time is malleable, fluctuates, and is unfolding. Figure 2 depicts one participant's trust evaluations across a twenty-minute segment. If the user's task ended at eleven minutes, their trust levels would be significantly different from those at 20 minutes. These findings about the variability of trust also suggest that dynamic trust measurements may be more important for different parts of a web-system. For example, our results indicated that user trust in information as a guide had the highest variability. Taking more frequent measurements for this specific trust assessment may be necessary to determine what interface elements caused such dramatic increases and decreases in trust.

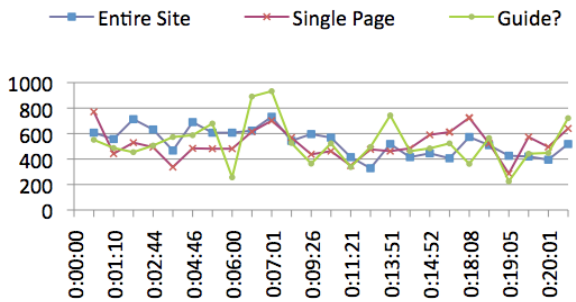

Figure 2. One participant's responses to trusting the entire website, a single webpage, or the information to guide them across one twenty minute interval.

Because participants were repeatedly polled for their trust values, they may have believed that their trust should therefore change over time. We acknowledge that continuously distracting the user from their primary goal can induce usability errors, thereby decreasing trust. However, as can be seen in Figure 2, the participant's responses were more than just a change over repeated polling. Participants responded to what was occurring in the interaction. Further, while trust values may actually be lower than if the participant had not been repeatedly polled, the graph in Figure 2 still demonstrates that a user's trust values differ by what is being assessed for trust (i.e., time 06:00).

The detection of a time effect across both information repositories indicates that trust may be more variable during the first five minutes of interaction. The TIME method facilitates detection of points where a user's trust changes in the interaction and suggests that measurement-taking frequencies may need to be increased to fully understand this important adjustment phase. The second finding from analyzing by time segments indicates that trust appears to decrease linearly from the beginning of the interaction

to the end. The average difference in trust measurements is a 12% decrease. This may indicate that when users first use a website they start with an initially high level of trust. As users continue though, frustration with the interface or other factors may impact the trust level reducing it to a 'normal' level. Designers of websites that desire high levels of trust should focus on this finding to create websites that sustain high levels of trust with interface elements that have been shown to increase trust (see [15, 31, 34] for examples). The main impact of the use of the TIME Method is the detection of these outcomes, which were previously obscured from analysis.

8. CONCLUSIONS & FUTURE WORK

In this paper we presented a review of different empirical methods to study trust in web systems. The benefits and drawbacks of each method were explored with quantitative methods providing numbers to associate with trust levels, and qualitative methods providing a way to identify likely determinants of trust. We then proposed the TIME Method of evaluation as a response to the drawbacks of previous methods. One of the primary findings we have been able to contribute to the body of work on trust in web systems is an understanding that trust is highly dynamic and a response to current interaction. Our second argument has been that there is a need to establish new methods, such as the TIME method, to examine constructs of trust.

To build on this work, we would like to validate the TIME Method against other online information repositories that do not have user-generated content (news sources like CNN.com), other kinds of web-systems (online communities like Facebook), and against and in combination with other trust evaluation methods (e.g., doing a pre-interview to identify trust facets and then using the form to poll them). The investigation of the relationship between trust and privacy, and how these two affect each other dynamically, should be explored using the TIME method. Example websites that could elicit privacy concerns are online forums and electronic medical records. Second we would like to do a comparison study of different evaluation methods, such as using Likert-type surveys along with open-ended interviews to assess the strengths and weaknesses of each model. We additionally would further evaluate variables believed to be causing an interaction effect. For instance, trust in the UCS was statistically lower than Wikipedia. Could a high base level of trust reduce the negative effects of poor usability or other factors on trust? And, if so, what is the cut off, and what factors lead to a "high enough" level of trust?

Work is also needed to adjust the evaluation methods of TIME. We used a period of approximately 30 seconds before presenting the form. We are unsure if this is the optimal amount of time, whether important information could be lost in the spaces between repeated measures, or if performing measurements after a click is appropriate. Different methods for presenting the form might also be evaluated. For instance, does a persistent form on the side of the window (rather than popping-up) allow for the same kind of interaction without being as disruptive? The important point is that the first instantiation of the TIME Method demonstrated that trust is variable. This indicates that further work should establish how to take these measurements with optimal results.

Last, more work is needed with respect to understanding the dynamic nature of trust, and constructs like it (e.g., quality, familiarity, experience). It is an assumption that the factors that decrease trust and credibility assessment will also increase or repair trust. In areas like eHealth and eCommerce, where understanding the minutia of a user's actions are critical, understanding trust and credibility and what causes it to change is paramount.

9. REFERENCES

[1] Amoroso, E., T. Nguyen, J. Weiss, J. Watson, P. Lapiska and T. Starr. Toward an approach to measuring software trust. in Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on. 1991.

[2] Bakker, M., R.T.A.J. Leenders, S.M. Gabby, J. Kratzer and J.M.L. Van Engelen, Is trust really social capital? Knowledge sharing in product development projects. Learning Organization, 2006. 13(6): p. 594-605.

[3] Beagrie, N., M. Doerr, M. Hedstrom, M. Jones, A. Kenney, C. Lupovici, et al., Trusted Digital Repositories: Attributes and Responsibilities. 2002, Research Libraries Group: Mountain View, CA. p. i-61.

[4] Bos , N., J. Olson, D. Gergle, G. Olson and Z. Wright. Effects of four computer-mediated communications channels on trust development. in Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves. 2002. Minneapolis, Minnesota, USA: ACM.

[5] Briggs, P., B. Burford, A. De Angeli and P. Lynch, Trust in online advice. Social Science Computer Review, 2002. 20(3): p. 321-321.

[6] Consolvo, S. and M. Walker, Using the experience sampling method to evaluate ubicomp applications. IEEE Pervasive Computing, 2003: p. 24-31.

[7] Deutsch, M., Trust and Suspicion. The Journal of Conflict Resolution, 1958. 2(4): p. 265-279.

[8] Dietz, G. and D.N.D. Hartog, Measuring trust inside organizations. Personnel Review, 2006. 35(5): p. 557-588.

[9] Ferrin, D., K. Dirks and P.P. Shah, Direct and Indirect Effects of Third-Party Relationships on Interpersonal Trust. Journal of Applied Psychology, 2006. 91(4): p. 870-883.

[10] Fisher, J., F. Burstein, K. Lynch and K. Lazarenko, "Usability plus usefulness = trust": an exploratory study of Australian health web sites. Internet Research, 2008. 18(5): p. 477-498.

[11] Flanagin, A.J. and M.J. Metzger, The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. New Media & Society, 2007. 9(2): p. 319-342.

[12] Fogel, J. and E. Nehmad, Internet social network communities: risk taking, trust, and privacy concerns. Computers in Human Behavior, 2009. 25(1): p. 153-60.

[13] Fogg, B.J. and H. Tseng, The Elements of Computer Credibility, in Conference on Human Factors in Computing Systems, Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit. 1999, ACM Press: Pittsburgh, Pennsylvania, United States. p. 80-87.

[14] Greenhalgh, T., G.W. Wood, T. Bratan, K. Stramer and S. Hinder, Patients' attitudes to the summary care record and HealthSpace: qualitative study. British Medical Journal, 2008. 336(7656): p. 1290-+.

[15] Hassanein, K.S. and M.M. Head. Building Trust Through Socially Rich Web Interfaces. in the Second Annual Conference on Privacy, Security, and Trust (PST'2004). 2004. Fredericton, Canada.

[16] Hovland, C., I.L. Janis and J.J. Kelley, Communication and Persuasion. 1953, New Haven, CT: Yale University Press.

[17] Huntington, P., D. Nicholas, B. Gunter, C. Russell, R. Withey and P. Polydoratou, Consumer trust in health information on the web. Aslib Proceedings, 2004. 56(6): p. 373-382.

[18] Iacono, C.S. and S. Weisband. Developing trust in virtual teams. in System Sciences, 1997, Proceedings of the Thirtieth Hawaii International Conference on. 1997.

[19] Intille, S.S., J. Rondoni, C. Kukla, I. Ancona and L. Bao. A context-aware experience sampling tool. in Conference on Human Factors in Computing Systems. 2003: ACM New York, NY, USA.

[20] Jones, S. and S. Marsh, Human-computer-human interaction: trust in CSCW. SIGCHI Bulletin, 1997. 29(3): p. 36-40.

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

[22] Khoo, K., P. Bolt, F.E. Babl, S. Jury and R.D. Goldman, Health information seeking by parents in the Internet age. Journal of Paediatrics and Child Health, 2008. 44(7-8): p. 419-423.

[23] Kittur, A., B. Suh and E.H. Chi. Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia. in Proceedings of the ACM 2008 conference on Computer supported cooperative work. 2008. San Diego, CA, USA: ACM.

[24] Larsson, M., A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery, 2009. 25(1): p. 14-20.

[25] Lewicki, R.J., E.C. Tomlinson and N. Gillespie, Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. Journal of Management., 2006. 32: p. 991-1022.

[26] Lumsden, J. and L. MacKay. How Does Personality Affect Trust in B2C e-Commerce? in Proceedings of the 8th International Conference on Electronic Commerce (ICEC'2006). 2006. Fredericton, New Brunswick, Canada.

[27] Maddux, W.W. and M.B. Brewer, Gender differences in the relational and collective bases for trust. Group Processes & Intergroup Relations, 2005. 8(2): p. 159-171.

[28] Marshall, L.A. and D. Williams, Health information: does quality count for the consumer? How consumers evaluate the quality of health information materials across a variety of media. Journal of Librarianship and Information Science, 2006. 38(3): p. 141-156.

[29] Mayer, R.C., J.H. Davis and D. Schoorman, An Integrative Model of Organizational Trust. The Academy of Management Review, 1995. 20(3): p. 709-734.

[30] Menon, A., A. Ddeshpande, M. PerriIii and G. Mzinkhan, Trust in online prescription drug information among Internet users: The impact on information search behavior after exposure to direct-to-consumer advertising. Health Marketing Quarterly, 2002. 20(1): p. 17-17.

[31] Norman, D.A., Emotional Design: Why we love (or hate) everyday things. 2004: Basic Books. 257.

[32] Riegelsberger, J., M.A. Sasse and J.D. McCarthy. Eye-Catcher of Blind Spot? The Effect of Photographs of Faces on eCommerce Sites. in Proceedings of the IFIP Conference on Towards The Knowledge Society: E-Commerce, E-Business, E-Government. 2002: Kluwer, BV Deventer, The Netherlands, The Netherlands.

[33] Rosenbaum, S.E., C. Glenton and J. Cracknell, User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library. Bmc Medical Informatics and Decision Making, 2008. 8.

[34] Rosenvinge, J.H., S. Laugerud and P. Hjortdahl, Trust in health Websites: A surevey among Norwegian Internet users. Journal of Telemedicine and Telecare, 2003. 9(3): p. 161-166.

[35] Semere, W., H.L. Karamanoukian, M. Levitt, T. Edwards, M. Murero, G. D'Ancona, et al., A pediatric surgery study: Parent usage of the Internet for medical information. Journal of Pediatric Surgery, 2003. 38(4): p. 560-564.

[36] Shiffman, S., Real-time self-report of momentary states in the natural environment: Computerized ecological momentary assessment. The science of self-report: Implications for research and practice, 2000: p. 277–296-277–296.

[37] Sillence, E., P. Briggs, L. Fishwick and P. Harris. Trust and mistrust of online health sites. in 2004 Conference on Human Factors in Computing Systems - Proceedings, CHI 2004, Apr 24-29 2004. 2004. New York, NY 10036-5701, United States: Association for Computing Machinery.

[38] Sillence, E., P. Briggs, P. Harris and L. Fishwick, Health Websites that people can trust - the case of hypertension. Interacting with Computers, 2007. 19(1): p. 32-42.

[39] Sillence, E., P. Briggs, P.R. Harris and L. Fishwick, How do patients evaluate and make use of online health information? Social Science & Medicine, 2007. 64(9): p. 1853-1862.

[40] Tio, J., A. LaCaze and W.N. Cottrell, Ascertaining consumer perspectives of medication information sources using a modified repertory grid technique. Pharmacy World & Science, 2007. 29(2): p. 73-80.

[41] Walther, J.B., Z.M. Wang and T. Loh. The effect of top-level domains and advertisements on health Web site credibility. 2004.

[42] Zahedi, F. and J. Song, Dynamics of trust revision: Using health infomediaries. Journal of Management Information Systems, 2008. 24(4): p. 225-248.

[43] Zaheer, A., B. McEvily and V. Perrone, Does Trust Matter? Exploring the Effects of Interorganizational and Interpersonal Trust on Performance. Organization Science, 1998. 9(2): p. 141-159.