Trust modelling for online transactions: A phishing scenario

Ponnurangam Kumaraguru
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
ponguru@cs.cmu.edu

Alessandro Acquisti
H. John Heinz III School of
Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213
acquisti@andrew.cmu.edu

Lorrie Faith Cranor
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
lorrie+@cs.cmu.edu

Abstract

Trust is an important component of online transactions. The increasing amount and sophistication of spam, phishing, and other semantic attacks increase users' uncertainty about the consequences of their actions and their distrust towards other online parties. In this paper, we highlight some key characteristics of a model that we are developing to represent and compare the online trust decision processes of "expert" and "non-expert" computer users. We also report on preliminary data we are gathering to validate, refine, and apply our model. This research is part of a broader project that aims at developing tools and training modules to help online users make good trust decisions.

1 Introduction

Trust is a precious asset in online transactions. Internet users often do not know whether to trust a certain online merchant with their personal information, or whether the email ostensibly from a legitimate company has rather been sent by an impostor. Making good online trust decisions is becoming increasingly difficult even for experts; it requires specialized knowledge (such as computer experience) and continually updated awareness of threats and attacks. It also requires the ability to deal with uncertainty and to properly assess risks. The increasing amount and sophistication of spam, phishing, and other semantic attacks are making trust decisions increasingly difficult. As a result, survey participants frequently mention lack of trust as one reason for not transacting online [19]. Internet novices and parents, in particular, refrain from taking advantage of several online features for fear of adverse consequences [45]. Because of trust issues, online users' privacy concerns also increase [4], [47], [48].

To inform the development of tools that will enable computer users to make better online trust decisions, we are surveying, representing, and contrasting "expert" and "non-expert" users' trust behaviors and decision processes. (The distinction between experts and non-experts we adopt is based upon an aggregate measure of the level of computer and Internet usage as well as knowledge of existing threats and possible defense strategies.) This research is part of a larger project that focuses on developing tools (such as new email applications) and training modules to improve trust decisions and protect Internet users against, specifically, "phishing" attacks.

In this paper, we focus on one component of such broader research agenda. We report on the key characteristics of a model that we are developing to represent and contrast the online decision processes of expert and non-expert users. We present some preliminary results from interviews and surveys we are administering to develop, validate, and apply our model.

The remainder of this paper is organized as follows: in the next section we present an overview of different trust models discussed in various research fields. In Section 3, we discuss the relation between phishing and trust. In Section 4, we introduce our modelling approach and explain the relationships between the model elements that can help us understand, represent, and contrast expert and non-expert decision processes. We also discuss representations of the model that help us in that goal. In Section 5, we present preliminary analysis of the data that we are collecting to evaluate the model in the phishing scenario. Finally, Section 6 sketches the conclusions from the analysis performed, as well as the limitations of the study and some future research directions.

2 Related literature

Trust is a concept with many dimensions [3], [10], [27], [33] that has been studied in many diverse disciplines: for instance, economics, where the focus is on agents' reputations and their effect on transactions [6], [7], [18], [20], [31], [34], [36]; marketing, where the focus is on strategies for consumers' persuasion and trust building [8], [9], [16], [44]; human computer interaction, where the focus is on the relation between design and usability of a system and users' reactions [10], [38], [39]; and psychology, where trust has been studied as an interpersonal and group phenomenon [40], [42].

Formal models of trust started appearing in the literature early in second half of the twentieth century. A common approach sees the trustor (the trusting party — for instance, an

individual buying something online) as considering engaging in an interaction with the trustee (the party to be trusted — for instance, an online merchant) [27]. The trustor engages in the transaction if and only if the level of trust is higher than his personal threshold trust value [46].

Researchers have identified several 'antecedents' of trust in their models. Antecedents are factors that affect trustors' trust level when the trustor is interacting with the trustee. Most of the models discussed in this paper focus on the trust level of the trustor towards the trustee. Mayer et al. have identified trustee's perceived integrity, benevolence, and ability as trust antecedents; they have shown that these antecedents may vary independent to each other [27]. Gefen evaluated Mayer et al.'s model using a survey instrument and showed that trust and trustworthiness should not be regarded as a single construct [17]. Lee et al. focused on three concepts — comprehensive information, shared values, and communication — as antecedents of trust. Their model also showed that transaction costs can negatively interact with trust in determining consumers' behavior [23]. The model developed by Mcknight et al. uses disposition to trust and prior trust in technology and the Internet (institution-based trust) as the antecedents for trust in online transactions. This model specifically shows the difference between trusting intentions and trusting behavior [28]. Tan et al. show that the trustors' trust in the transaction can be split into trust on the trustee (party trust) and trust in the control mechanism (control trust) of the transaction [46]. Ang et al. focus on the trustees' ability to deliver on its promises, the trustees' willingness to rectify any problems arising from dissatisfaction, and trustees' respect towards the personal privacy of the trustor [1]. Bhattacherjee has highlighted familiarity (the trustors' prior interactions and experiences with the trustee) as one of the antecedent to trust [5]. His model takes integrity, benevolence, and ability from Mayer's model as antecedents to trust. Corritore et al. created a model that focuses on perceptual factors — credibility, ease of use and risk - as the antecedents [10]. Egger focuses on disposition to trust, prior knowledge or experience, information and attitudes transferred from others, trustee's reputation, organization involved, and trust on information technologies and the Internet in general as antecedents in defining the trust model [15]. Riegelsberger et al. developed a trust model that focuses on the incentives for trustworthy behavior. They showed that contextual properties (motivation based on temporal, social, and institutional embeddedness) and the trustors' intrinsic properties (ability, motivation, and benevolence) are antecedents for trustworthy behavior [37], [38], [39]. Antecedents in various models are highlighted in Table 1.

Researchers have also worked on formalizing trust with computational models [26], [30], [35]. These trust models have been used to build trustworthy agents [32].

Table 1. *Online Trust Antecedents (arranged in chronological order)*

Authors	Antecedent factors
Mayer et al. [27]	integrity, benevolence, and ability
Lee et al. [23]	comprehensive information, shared value, and communication
Mcknight et al. [28]	disposition to trust, trust in technology and the Internet (institution-based trust)
Tan et al. [46]	party trust and control trust
Ang et al. [1]	ability to deliver, willingness to rectify, and personal privacy
Bhattacherjee [5]	integrity, benevolence, ability, and familiarity
Corritore et al. [10]	perceptual factors — credibility, ease of use, and risk
Egger [15]	disposition to trust, prior knowledge or experience, information from others, trustee's reputation, and trust on information technologies
Riegelsberger et al. [39]	temporal/social/institutional ability, benevolence, and motivation

3 Phishing and trust

Phishing is a semantic attack which targets the user rather than the computer. Phishing attacks take advantage of the way we assign meaning to content: phishers are successful by making use of the differences between the system model and the mental model [29], [41]. Typically, such attacks aim at misleading the victim into willingly revealing his personal information. Phishing emails, for instance, appear to come from legitimate and trustworthy sources. Phishing is rapidly growing and is cause for significant concern among users and IT professionals. Over 20,000 unique phishing attacks were reported to the Anti-Phishing Working Group (APWG) in May 2006, compared to over 17,000 attacks in April 2006 [2].[1]

Phishers trick users into trusting untrustworthy websites and revealing their personal information. At the same time, the increasing sophistication of these attacks makes them hardly distinguishable from legitimate emails, and reduces the trust users afford to genuine sites. Both the damaging consequences of false negative (not recognizing a phishing attack as such, and therefore providing personal information to criminals) and the costs of false positive (disregarding a genuine communication as an attack, and therefore failing to act upon a certain required action, which may cause opportunity or actual costs) can be significant.

Previous trust research offers little guidance on how these online trust decisions could be made easier. Previous models have mostly focused on understanding the effect of trust on transactions (the economic approach); developing

[1] For some recent works on phishing, see [12], [14], [22], and [25].

means to persuade consumers to transact online (the marketing approach); and understanding the relation between a system's usability and the trust behavior of the online users (the HCI approach). Our goal is to educate, support, and improve the trustor's online decisions — which may at times involve assigning more or less trust, depending on the context.

Our research is closest to those models in the literatures that have identified relationships between trust and factors such as a trustee's reputation or benevolence, and a trustor's ability or knowledge. However, since online trust issues (and in particular, phishing problems) arise from the dichotomy between the system model and the user's mental model, we depart from that research in that our approach is informed by the game theoretical literature on signalling [43] and the relation between uncertainty, expected outcomes, and users' decision making. Recent research has, in fact, shown that people fall for phishing attacks because of misjudgement of risks, misinterpretation of available information (what we will refer to as "signals" in Section 4), and erroneous knowledge about security. [11], [12], [21], [49], [50]. Even more than in offline scenarios, online signals that inform trust decisions can be easily spoofed (for example, it is easier to spoof a Barnes and Nobles website than to spoof a physical store of the size of Barnes and Nobles) [22], [24]).

In the rest of this paper, we highlight some key characteristics of a model that we are developing to represent and contrast the online decision processes of "expert" and "non-expert" computer users, and we report on preliminary data we have gathered to design, validate, and apply our model.

4 Trust modelling

At the root of online trust issues (from phishing to privacy violations or other online scams) are two factors: the dichotomy between the signals that users can observe about the underlying state of key variables of importance to their decision process; and malicious entities that manipulate to their advantage those signals, the associated underlying variables, and, in turn, the user's decision process and well-being. Our goal is to understand how users take advantage of available information (signals that may or may not be accurate) to infer and evaluate possible consequences and states of variables that will affect their well-being.

The dichotomy between signals and true states is a form of incomplete information that is also studied in game theoretical signaling models. Although we do not adhere to the rational and common knowledge assumptions of those models to represent actual users' behavior, we find some of the tools developed in that area useful to represent trust decision processes. For example, the observed "sent" field in an email is a signal for the sender of that email — however, that signal may have been spoofed by an attacker.

Specifically, we represent trust decision problems as combinations of the following stylized components:

- *States of the world* are the true realizations of the variables that affect, primarily, the user's well-being, but secondarily also her decision process. Examples of states of the world are whether a certain email was actually sent by a colleague or by a spammer, or whether the CitiBank page a user just accessed is a legitimate page residing on the bank's servers or is a fake on a third-party malicious host.
- *Signals* represent the information available to users about the states of the world. They are, in other words, noisy functions of the underlying states of the world that may be more or less informative. Examples of signals are the "from" field in an email, that provides sender information, or the URL address on the top of a browser.
- *Actions* are the set of things that a user may do in a certain scenario. For example, a user that receives an email from (ostensibly) a colleague, but with a suspicious attachment, may open it, delete it, scan it with an anti-virus software, ignore it, and so on.
- *Decisions* refers to the adoption of a specific (set of) actions, as determined by personal heuristic or rational deliberation informed by the evidence available to users about the states of the world and the consequences of the user's actions, through which the user attempts to attain some objective measure of well-being. An example of decision may be: when the sender is unknown, delete her email.
- *Well-being* is a measure of the user's satisfaction that depends on combinations of certain states of the world and the user's decisions. For example, a user's well-being will be positively affected when the user ignores a suspicious email if it was actually sent by an impostor, and negatively affected when the user fails to do so.
- *Attackers* are entities that can deliberately influence the signals, the states of the world, and therefore the user's decision to their own advantage. Some examples of attackers are spammers or phishers.

We define online trust problems as those that arise when dichotomies between signals and underlying states can affect the user's decisions and well-being, and when attackers can affect signals, states, and decision processes. In other words our goal is to understand users' decision process in this context: users make **decisions** among alternative **actions**, in order to satisfy certain personal **well-being** objectives. Such decisions are informed by noisy **signals** about the true underlying **states of the world**, and external **attackers** can affect those world states, those signals, and therefore the users' decisions, and, ultimately, well-being. For instance, in a phishing scenario, the user may receive an email from what looks like a legitimate sender **[signal]**. The email may have been actually sent by a scammer **[attacker]** and the linked site the user is referred to may be a phishing site **[state of the world]**. The user can follow or not follow the email instructions **[actions]**, and will do so depending on personal heuristics, knowledge, and expected consequences associated with dealing or not with those instructions **[decision]**.

4.1 Relationships between elements

Using symbols allows us to compactly show the functional relations between the various modelling components.

For simplicity, we will call the states of the world W, the signals S, the actions A, the decisions D, the attackers X, and the user's well-being U.

Signals may incorporate accurate or inaccurate information about the actual states of the world, and states of the world are observed through those signals:

$$W \rightarrow S \tag{1}$$

Decision processes represent the selection of actions resulting from the consideration of the well-being associated with combinations of actions and states of the world. Decisions are therefore affected by signals (that inform the user's expectation about his well-being and various states of the world) but also directly by some states of the world (for instance, how tired a user is may affect how the user decides about how to act on a certain email):

$$S, W \rightarrow A \tag{2}$$

Actual well-being is determined by the combination of the user's actions and the states of the world:

$$W, A \rightarrow U \tag{3}$$

Phishers can influence the states of the world (for instance, phisher can send an email from a known sender account to the user), but also the signals available to users (for instance, phishers can modify the URL presented in the email to the user).

$$X \rightarrow S, W \tag{4}$$

This representation captures just some of the main functional relations that inform trust decision, and shows how attackers can end up affecting the user's well-being:

$$X \rightarrow S \rightarrow A \rightarrow U \tag{5}$$

4.2 Representations

The above elements can be useful to represent user's decision processes in trust scenarios. Figures from 1 to 4 are representations of trust decisions informed by the above approach. They are being populated with the results of our ongoing interviews and surveys (see Section 5.1). They highlight the relationships between signals and the states of the world for a user and the role and accuracy of various signals in a user's decision process. By populating and then contrasting models from the expert and non-expert points of views we aim at learning critical insights to build systems that help non-experts make better trust decisions.

Figure 1 should be interpreted as having three layers rather than as a Venn diagram. The three layers are: signals, states that affect the decision, and states that affect the well-being of the users. Signals represent the information available to and used by subject in her decision process. Signals can be more or less informative about the true states of the underlying variables — in other words, signals could be correct or incorrect. They are represented as a layer above the two "states" layers. Both "states" layers refer to actual states of the world. However, one layer represents states that may affect the user's well being (for instance, whether an email contains a virus or not). The other layer represents states that may affect the user's decision process — this could happen outside of an actual deliberative process (for instance, how tired or impressionable the user is may affect her decision process in dealing with potential phishing attacks); or indirectly, through signals on which the user bases her decision.

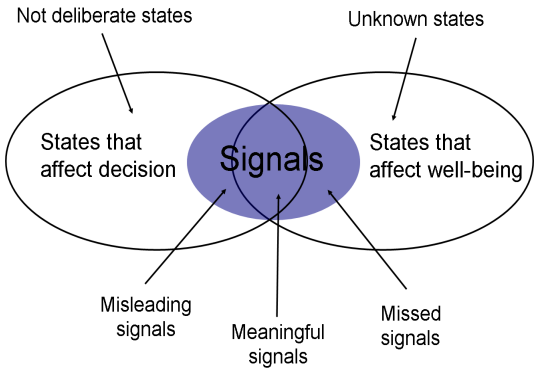

Figure 1. Trust Model

The space in which signals, states that affect well-being, and states that affect decision overlap, represents information that enters the user's decision process and is about states of the world that may affect the user's well being. The signals that lie at the overlap of the two "states" sets are therefore *meaningful signals*, in the sense that they carry information about the underlying states of variables that directly affect the user's well being, and also enter the user's decision process. For instance, an individual may use an email's subject and sender information as signals to decide whether to open or delete it; those signals are possibly informative about the identity of the true sender of the email, and therefore relate to states of the world that may be relevant to the user and ultimately affect her well-being.

Certain pieces of information may affect the user's decision process without being truly relevant to states that will affect the user's well being. In Figure 1 they are called *misleading signals*. For instance, a user may rely on a wrong or uninformative type of signal — e.g., the font size of the text included in an email — to determine whether an email is legitimate.

Signals that are related to underlying states of the world that may affect the user's well being but are ignored or not considered by her are represented as *missed signals* in Figure 1. For instance, users may not be aware of the signals that show the route or path taken by an email but that are informative about its actual sender.

Certain states of the world may affect the user's well-being but are not known to the user through the signals — in Figure 1 they are represented as *unknown states of the world*. For instance, notwithstanding all the information the user can seek, she still may not be able to discern who is the

actual sender of a certain email.

Finally, certain states of the world affects the decision process directly, without an actual deliberative process, rather than through signals: they are represented as *not deliberate states of the world* in Figure 1. For instance, tiredness may trigger careless behavior in an otherwise usually careful user.

One of the goals of our research is to populate Figure 1 with the signals used by experts and non-experts and contrast the relative sizes of the various layers. We would expect the experts' version of that representation to look like Figure 2, in which one can detect a significant overlap of the states that affect the decision process and states that affect the well-being (experts make significant use of meaningful signals, and have fewer missed or misleading signals than non-experts).

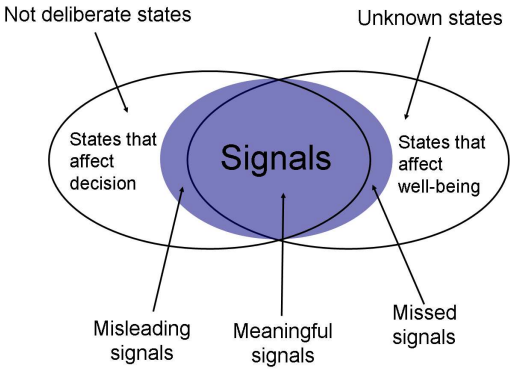

Figure 2. Trust Model — Experts

The non-experts' version may look like Figure 3: we can observe a small overlap of the states that affect users' decision and the states that affect their well-being. This in turn creates a small space for meaningful signals and a large space for missed and misleading signals. In comparison to the experts the size of the missed and misleading signals is larger for the non-experts.

A different representation for the experts and non-experts' decision processes that we are developing based on the results of our interviews and surveys is presented in Figure 4. There, we plot the *accuracy* of the signals adopted by a user in her decision process against the *weight* assigned by the user to that signal in her decision. Accuracy refers to how informative a certain type of information is considered to be, *vis a vis* the true states of the underlying variables. For instance, reading the text links in email may provide less accurate information than studying the actual URLs in the html code of the email. The base line signal accuracy is obtained from interviews with experts, to be contrasted to the data gathered from interviews with non-experts. The weights that users assign to various signals in their decision processes are also being extracted from our interviews. For instance, experts may put more weight on URLs listed in an email ostensibly sent by PayPal than on the sender information, because they know that the latter is more easily spoofable. In general, experts should give more weight than non-experts to types of signals that are more accurate and informative. In Figure 4, this is represented by the contrast between bullets (representing non-experts data) and triangles (representing experts data).

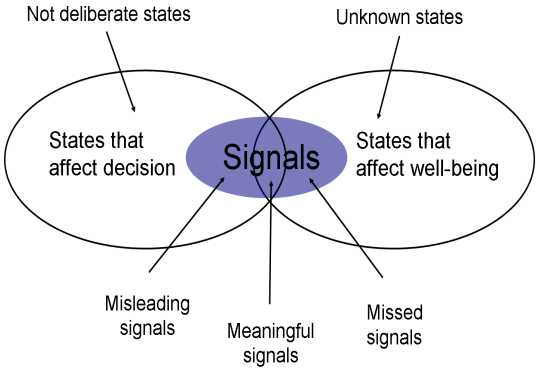

Figure 3. Trust Model — Non-Experts

5 The phishing scenario

In this section, we apply our model and representations to the phishing scenario. We present a preliminary analysis from our ongoing interviews. This analysis should only be considered as a work-in-progress, illustrative of the links and comparisons we are trying to establish.

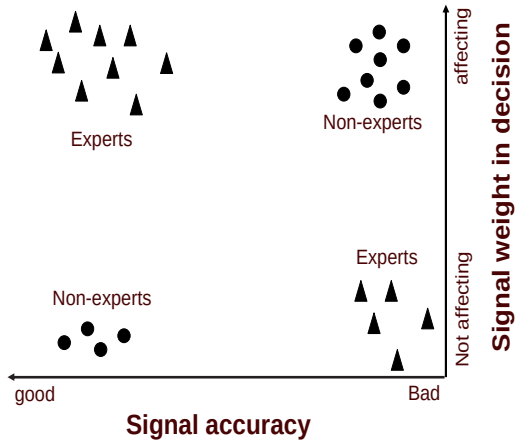

Figure 4. Trust Model Graph

5.1 Preliminary data analysis

We have been conducting interviews with both computer security experts and non-experts about their knowledge, sensitivity, and behavior related to online transactions and online information security and privacy. (The application by other members of our team of a "mental models" approach to a similar data set is presented in [13]).

During the interviews we asked the subjects about their decision making strategies in the following scenarios: (1) receiving emails, (2) accessing websites, (3) downloading software, and (4) buying products online. Interview data are being used to create and refine the model presented in Section 4.

5.1.1 Data Set

Our data set comprises interviews with twenty four non-expert participants, two experts, and an additional group-interview with experts. Participants were recruited by posting paper flyers and sending emails to groups in and around Carnegie Mellon University (CMU). Individuals doing research in the area of security and privacy at CMU were considered experts. Non-experts were people who replied to our fliers and answered "no" to three screening questions: 1) whether they had ever changed preferences or settings in their web browser, 2) whether they had ever created a web page, and 3) whether they had ever helped someone fix a computer problem. If somebody had said "yes" to any of the above questions, more details were requested, and subjects were excluded if they had any experience with security [13].

5.1.2 Meaningful signals

Signals that enter the user's decision process and are associated with states that can affect her well-being are represented as "meaningful signals" in Figure 1. Through the interviews we found that the number of these types of signals that non-experts consider is limited; from Table 2 we can see that non-experts listed fewer signals than experts. Lack of computer system knowledge and lack of knowledge of security and security indicators, as described by Dhamija et al. [12], may explain this dichotomy.

A typical response of a non-expert regarding the signals used while making a decision about how to deal with an incoming email is: "who it is coming from, if I recognize the email address will open it and if it is something I am expecting will open it." A typical response of experts is: "... I look for the person from whom it is coming, subject information, length or size of the email, whether there is an attachment and sometimes route and header information." Experts also use different techniques to choose their course of action; one of the experts mentioned " ... important emails have tags on them. They [mails] are expected to have a tag on the email address to me but I will see the tag or it just goes out. [email server] allows a username + and a tag @ the host and whatever comes after the plus sign is ignored by the mailer so it passes as a tag ... [which can be used to] sort emails that way." Experts felt that some meaningful signals are not really useful in making a decision. For example one expert said, "... what I figure is the SSL stuff tells you whether your data is protected in transit but it tells you nothing about what they are actually going to do once they get it. And there is the risk of it being intercepted in transit is relatively low compared with it being intercepted elsewhere ... I don't feel like I have any way of really knowing if it is actually secure, I don't think the SSL thing really tells you a whole lot." The experts also feel that absolute security is impossible: "I think it's [absolute security] a relative thing. It's a spectrum of being secure and insecure and I think the point absolute security that will satisfy [everybody] is probably impossible ...".

We also asked experts and non-experts about the signals they consider when making decision about downloading software after being prompted by a website (for example, a website prompting the user to download Macromedia Flash to view its contents). A typical response from a non-expert was: "look at the site and if it is actually something that I want to read, if I recognize the name of the site and the program to download." Whereas one of the experts mentioned the following signals: "explanation provided on the need for the software; do I really want it? can I find it from reputable source like www.download.com? what is the extension type of the file? sometime read user comments on websites about the software." In addition, one of the experts mentioned that if the software is recommended by other websites then they might decide to download and use it.

When experts were asked about what can be done to help non-experts make use of the information available to them, one of the suggestions was not to bury useful information (such as header information and route information) in the options of the email client, which are difficult to find. Experts feel that it will be useful if this information is made available to users easily and in a usable form. Experts also find it difficult to say things to the non-experts to support their decision making. This can be seen from one expert's response: "... when I try myself to describe this [tell non-experts about things to look for] to my friends and family, I find it very difficult, I try to warn them like these kinds of scams exist and then therefore one should be careful in doing their finances, other than that I really I am not sure how to describe it".

5.1.3 Missed signals

Useful information ignored by users is defined as "missed signals" in our model representation. From Table 2 we can see that none of the following signals in the email scenario were mentioned by non-experts: length or size of the email, route information, "whois" information of the link in the email, header information, etc. However, most of these signals were frequently mentioned by experts. In addition, the following signals were not mentioned in the website scenario: HTTPS, broken images, status bar, SSL, "whois," various toolbars, etc. Non-experts typically make decisions in the email situation based only on the sender's email address. Non-experts are often not aware of many other signals which can be more useful than the ones they use. We can also observe that they are not aware that most of the signals they mentioned can be spoofed. One implication is that non-experts should be made aware of the problems associated with some of the signals they use and the availability of other signals. One way of improving non-experts'

Table 2. *Signals in emails and websites*

	Signals	
	Non-experts	Experts
Emails	Sender	Sender
	Reputation of the sender	Reputation of the sender
	Subject	Subject
		Length or size
		Route information
		Language of the content
		whois
		Header information
	Spam filter	Spam filter
Websites		HTTPS
	Security lock	Security lock
	Reputation of the website	Reputation of the website
		Broken images
	Address bar	Address bar
		Status bar
		SSL
		whois
		Toolbars

decision process may be to increase their awareness of attacks, means of protection, and signals that reveal something about the true nature of an email or a website. Non-experts should be informed about basic Internet safety rules such as: "don't believe everything you read;" "a polished appearance is not the same as substance;" "if something is too good to be true, it probably is [24]."

5.1.4 Misleading signals

Certain pieces of information may affect the user's decision process although they do not really relate to the underlying state of variables that influence the user's well-being; these are "misleading signals." Signals that are meaningful sometimes turn to be misleading for non-expert users. For example the following signals are used by non-experts to make their decision: "professionalism of the content in the website" and "reputation and brand of the website." Non-experts are not aware that all these signals can be easily spoofed.

In the data that we have collected, we also see that some of the signals discussed in Table 2 can be grouped into the antecedents discussed in Section 2. For example "reputation of the sender" and "reputation of the website" from our data is comparable to "trustee's reputation" by Egger et al. [15].

6 Conclusions

We have presented our ongoing efforts at modelling online trust decisions, focusing on a phishing scenario. Our goal is to learn about, represent, and contrast expert and non-expert users' decision processes. Our approach is informed by the signaling and game theoretical literature but is grounded in ongoing empirical observation. We presented a generic model representing trust and used the generic model to populate data for the phishing scenario. We also compared and contrasted data from experts and non-experts. In this paper, we also presented preliminary evidence from the data that we have been collecting through interviews with experts and non-experts.

This research is part of a broader project that has the ultimate goal of developing tools and training modules to help online users make good trust decisions.

Our results are preliminary and will develop with the rest of our broader agenda. For instance, experts and non-experts interviews are being followed by surveys with a significantly larger number of users. Still, while we cannot yet extrapolate generalizable results from this data, we are gaining glimpses at the differences between the experts' and non-experts' online trust decision making. This could help us in the design and implementation of tools to aid non-experts in making better online trust decisions.

Acknowledgements

We gratefully acknowledge support from National Science Foundation grant number 0524189 entitled "Supporting Trust Decisions," and from the Army Research Office grant number DAAD19-02-1-0389 entitled "Perpetually Available and Secure Information Systems." The authors would like to thank all members of the Supporting Trust Decisions project for their feedback. In particular, the authors would like to thank Julie Downs and Mandy Holbrook for their comments and discussions. We would also like to thank José Brustoloni, Sven Dietrich, Jason Hong, Norman Sadeh, Serge Egelman, Ian Fette, Mark Huneke, Faisal Jawdat, and Steve Sheng for their thoughtful comments.

References

[1] L. Ang, C. Dubelaar, and B. C. Lee. To trust or not to trust? a model of internet trust from the customer's point of view. In *Proceedings, 14th Bled Electronic Commerce Conference.*, pages 25–26, June 2001.

[2] Anti-Phishing Working Group. Phishing activity trends report. Technical report, Anti-Phishing Working Group. Retrieved Aug 11, 2006, http://www.antiphishing.org/reports/apwg_report_May2006.pdf.

[3] I. Araujo and I. Araujo. Developing trust in internet commerce. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, 2003. Retrieved Sep 13, 2005, http://portal.acm.org/citation.cfm?id=961324.

[4] S. Bellman, E. J. Johnson, S. J. Kobrin, and G. L. Lohse. International Differences in Information privacy concerns: A global survey of consumers. *The Information Society*, 20:pp. 313 – 324., 2004.

[5] A. Bhattacherjee. Individual Trust in Online Firms: Scale Development and Initial Test. *Journal of Management Information Systems*, 19(1):211–242, Summer 2002.

[6] L. M. B. Cabral. The Economics of Trust and Reputation: A Primer. Technical report, New York University and CEPR, May 2002. Retrieved Feb 20, 2006,

http://pages.stern.nyu.edu/ lcabral/reputation/ Reputation_June05.pdf.

[7] J. Cave. The economics of cyber trust between cyber partners. Retrieved Feb 20, 2005, http://www.foresight.gov.uk/Previous _Projects/Cyber_Trust_and_Crime_Prevention/Reports_ and_Publications/Economics_of_Trust_ Between_Cyber_Partners/Economics.pdf.

[8] A. Cavoukian and T. Hamilton. *The Privacy Payoff, How Successful Business Build Consumer Trust.* McGraw-Hill Ryerson Trade., 2002.

[9] R. K. Chellappa and R. Sin. Personalization versus Privacy: An Empirical Examination of the Online Consumers Dilemma. Vol. 6, No. 2-3, 2005. Retrieved Sep 13, 2005, http://asura.usc.edu/ ram/rcf-papers/per-priv-itm.pdf.

[10] C. L. Corritore, B. Kracher, and S. Wiedenbeck. On-line trust: concepts, evolving themes, a model. *Int. J. Hum.-Comput. Stud.*, 58(6):737–758, 2003.

[11] R. Dhamija and J. Tygar. The Battle Against Phishing: Dynamic Security Skins. *Symposium On Usable Privacy and Security*, 2005. Retrieved Feb 10, 2006, http://cups.cs.cmu.edu/soups/2005proceedings/p77-dhamija.pdf.

[12] R. Dhamija, J. D. Tygar, and M. Hearst. Why Phishing Works. *To appear in the Proceedings of the Conference on Human Factors in Computing Systems (CHI2006), 2006*, 2006. Retrieved Feb 10, 2006, http://www.sims.berkeley.edu/ rachna/papers/ why_phishing_works.pdf.

[13] J. Downs, M. Holbrook, and L. Cranor. Decision Strategies and Susceptibility to Phishing. In Proceedings of the 2006 Symposium On Usable Privacy and Security, 12 - 14 July, 2006, Pittsburgh, PA.

[14] C. E. Drake, J. J. Oliver, and E. J. Koontz. Anatomy of a phishing email. Technical report, MailFrontier. Retrieved Feb 27, 2006, http://www.mailfrontier.com/docs/MF_Phish_ Anatomy.pdf.

[15] F. Egger. *From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce.* PhD thesis, Eindhoven University of Technology (The Netherlands), http://www.ecommuse.com/research/publications/ thesis.htm, 2003.

[16] B. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do.* Morgan Kaufmann., December, 2002.

[17] D. Gefen. Reflections on the dimensions of trust and trustworthiness among online consumers. *SIGMIS Database*, 33(3):38–53, 2002.

[18] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, and C. L. Soutter. Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

[19] S. Grabner and E. A. Kaluscha. Empirical research in online trust: a review and critical assessment. *Int. J. Hum.-Comput. Stud.*, 58(6):783–812, 2003.

[20] G. A. Guerra, D. J. Zizzo, W. H. Dutton, and M. Peltu. Economics of Trust in the Information Economy: Issues of Identity, Privacy and Security. Technical report, Oxford Internet Institute, April 2003. Retrieved Feb 20, 2006, http://www.oii.ox.ac.uk/resources/publications/RR1.pdf.

[21] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social Phishing. Retrieved March 7, 2006, http://www.indiana.edu/ phishing/social-network-experiment/phishing-preprint.pdf.

[22] L. James. *Phishing Exposed*. Syngress, November 10, 2005.

[23] J. Lee, J. Kim, and J. Y. Moon. What makes internet users visit cyber stores again? key design factors for customer loyalty. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 305–312, New York, NY, USA, 2000. ACM Press.

[24] R. Lininger and R. D. Vines. *Phishing: Cutting the Identity Theft Line*. Wiley, publishing Inc., 2005.

[25] Mail Frontier. Mailfrontier field guide to phishing. Retrieved Jan 28, 2006, http://www.mailfrontier.com/docs/field_guide.pdf.

[26] S. P. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, 1994. citeseer.ist.psu.edu/marsh94formalising.html.

[27] R. C. Mayer, J. H. Davis, and D. F. Schoorman. An integrative model of organizational trust. *The Academy of Management Review.*, 1995(3):709–734, July, 1995.

[28] D. H. Mcknight, V. Choudhury, and C. Kacmar. Trust in e-commerce vendors: a two-stage model. In *Proceedings of the twenty first international conference on Information systems*, 2000. Retrieved Sep 13, 2005, http://portal.acm.org/citation.cfm?id=359640.359807.

[29] R. C. Miller and M. Wu. Fighting Phishing at the User Interface. Aug, 2005. In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[30] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002.

[31] D. C. Mutz. Social trust and e-commerce, experimental evidence for the effects of social trust on individuals' economic behavior. *Public Opinion Quarterly*, 69(3):393–416, 2005. Retrieved Feb 20, 2006, http://poq.oxfordjournals.org/cgi/reprint/69/3/393.

[32] A. S. Patrick. Building trustworthy software agents. 6(6):46–53., 18-19 October 2002. Retrieved Sep 13, 2005, http://www.andrewpatrick.ca/cv/building_trustworthy _agents.pdf .

[33] A. S. Patrick, P. Briggs, and S. Marsh. Designing Systems That People Will Trust. Aug, 2005. In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[34] M. G. Pollitt. The economics of trust, norms and networks. *Business Ethics - A European Review*, 11(2):119–128, 2002. Retrieved Feb 20, 2006, http://www.electricitypolicy.org.uk/people/pollitt/ economicstrust.pdf.

[35] A. A. Rahman and S. Hailes. A distributed trust model. In *NSPW '97: Proceedings of the 1997 workshop on New security paradigms*, pages 48–60, New York, NY, USA, 1997. ACM Press.

[36] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebays reputation system. *Draft Version for review by NBER workshop*, 2001. Retrieved Feb 20, 2006, http://www.si.umich.edu/ presnick/papers/ebayNBER/RZNBERBodegaBay.pdf.

[37] J. Riegelsberger and M. A. Sasse. Trustbuilders and Trustbusters: The Role of Trust Cues in Interfaces to e-Commerce Applications. Oct 3–5, 2001.

[38] J. Riegelsberger, M. A. Sasse, and J. D. mccarthy. Shiny happy people building trust?: photos on e-commerce websites and consumer trust. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003. Retrieved Sep 13, 2005, http://portal.acm.org/citation.cfm?id=642634&coll=ACM &dl=ACM&CFID=53973409&CFTOKEN=34372821#.

[39] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. The Mechanics of Trust: A Framework for Research and Design. *International Journal of Human-Computer Studies*, 62(3):381–422, 2005.

[40] P. Salovey and A. Rothman. *Social Psychology of Health*. Psychology press, 2003.

[41] B. Schneier. Are you sophisticated enough to recognize an internet scam? Retrieved Feb 24, 2006, http://www.schneier.com/essay-035.html.

[42] C. L. Scott. Interpersonal trust: A comparison of attitudinal and situational factors. *Human Relations*, 33(11):805–812, 1980. http://hum.sagepub.com/cgi/reprint/33/11/805.

[43] A. M. Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

[44] J. Stanford, E. R. Tauber, B. Fogg, and L. Marable. Experts vs. Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites. 2002. Retrieved Sep 13, 2005, http://www.consumerwebwatch.org/dynamic/web-credibility-reports-experts-vs-online.cfm.

[45] Susannah Fox et al. Trust and privacy online: Why Americans want to rewrite the rules. August 20, 2000. Retrieved Sep 13, 2005, http://www.pewinternet.org/pdfs/PIP_Trust_Privacy_Report.pdf.

[46] Y. H. Tan and W. Thoen. An Outline of a Trust Model for Electronic Commerce. 14(8), 2000.

[47] A. Westin and Harris Louis & Associates. Health Information Privacy Survey. 1993. Conducted for Equifax Inc. 1,000 adults of the national public.

[48] A. Westin and H. Interactive. IBM-Harris Multi- National Consumer Privacy Survey for IBM. Approximately 5,000 adults of the U.S. Britain and Germany. Technical report, 1999.

[49] M. Wu. *Fighting Phishing at the User Interface*. PhD thesis, MIT, 2004. Retrieved Feb 10, 2006, http://groups.csail.mit.edu/uid/projects/phishing/proposal.pdf.

[50] M. Wu, R. C. Miller, and S. L. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? *To appear in the Conference on Human Factors in Computing Systems (CHI 2006)*, 2006. Retrieved Feb 10, 2006, http://www.simson.net/ref/2006/CHI-security-toolbar-final.pdf.