

Trust at First Sight? A Test of Users' Ability to Identify Trustworthy E-commerce Sites

Jens Riegelsberger, M Angela Sasse & John D McCarthy

*Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK*

Tel: +44 20 7679 3643

Fax: +44 20 7387 1397

Email: {j.riegelsberger, a.sasse, j.mccarthy}@cs.ucl.ac.uk

Consumer trust in e-commerce is a key concern in current HCI research. In this paper, we report a study that investigated (1) if users can correctly identify trustworthy vendors based on cues they perceive in the interface. A further aim was to test (2) if users' judgement can be influenced by the introduction of an affective element — an employee photo. Since such elements have been reported to have negative effects on usability, we also checked for (3) effects of the photo on users' visual gaze pattern and task performance when interacting with the sites. At first sight of a page, users' ability to identify trustworthy vendors was not better than chance. Only after detailed exploration could users reach correct trust decisions. A photo only had an effect on participants' first impression of a vendor. We did not find effects of the photo on task performance.

Keywords: e-commerce, Web design, consumer trust, surface cues, interpersonal cues, eye-tracking, affective trust, decision-making, photos.

1 Introduction

There is wide agreement that e-commerce currently suffers from a lack of consumer trust [Brynjolfsson & Smith 2000; Einwiller 2001; Egger 2002]. However, there is little knowledge about what individual online vendors can do to overcome this 'lack of trust' and show that they are professional, reliable and reputable — in short

— trustworthy. For traditional shops, the situation is different: Most consumers have a high level of generalized trust, based on previous experience, institutional frameworks, and branding. In most situations, they will not consider fundamental problems such as being defrauded or not receiving a good at all. Their risk considerations will mainly focus on the quality or pleasantness of the service. With regard to these risks, they can make quick trust assessments based on cues they perceive from the interior, the goods and the sales staff [Einwiller 2001]. In fact, most purchases are routine actions, where individual trust reasoning will be replaced by an *expectation of continuity* [cf. Luhmann 1979]. Not so on the Web: Users perceive many risks and are unsure to what extent they can base trust decisions on the cues they perceive through a vendor's interface (*surface cues*). In our recent qualitative study, some users voiced concerns that “anyone could knock up a trustworthy-looking Web site” [Riegelsberger & Sasse 2001]. Thus, the first aim of this study is to find out (1) *how well shoppers can differentiate trustworthy, i.e. reliable and professional vendors, from untrustworthy ones, based on surface cues.*

Several HCI researchers have looked at how e-commerce vendors can optimize such surface cues [Studio Archetype & Cheskin Research 1999; Shneiderman 2000; Riegelsberger & Sasse 2001; Nielsen Norman Group 2002; Egger 2002]. This study puts one cue from these guidelines on *designing for trust* to the test: the ‘friendly face’. We chose this cue of interpersonal interaction because it is commonly used in advertising to convey an emotional atmosphere and engender trust. Photos of employees, virtual shopping assistants, or live text-based assistance can also be seen on several e-commerce sites now. Not surprisingly, such cues and their effect on online trust have recently received attention in the HCI community (see Section 2.2). Hence, the second aim is to investigate (2) *if the judgement of shoppers can be influenced by adding a photo of an employee to the interface of an existing online vendor.*

However, adding elements whose sole purpose is to communicate trustworthiness might have adverse effects on usability. When an online shopper's primary task is to select a product from a list, additional elements might make search more difficult. Hence, we want to investigate (3) *the effect of an employee photo on visual attention and usability.* To identify fine-grained effects on visual search patterns, we use eye-tracking as a method.

The paper starts by introducing work on trust and affective cues in the HCI and marketing literature (Section 2). Based on this we describe how the experimental design (Section 3) addressed the research questions introduced above. In Section 4, we present the results and discuss their implications in Section 5. We finish with substantive (6.1) and methodological conclusions (6.2).

2 Background

2.1 Consumer Trust in E-commerce

Trust has been described as willingness to be vulnerable based on positive expectations about the actions of others [Rousseau et al. 1998; Corritore et al. 2003]. This implies that trust is only required in the presence of *risk* and *uncertainty*. Online shopping carries more risk than off-line shopping because it relies on a

complex socio-technical system that stretches interaction over space and time [Brynjolfsson & Smith 2000], a process that is called *dis-embedding* [Giddens 1990]. Dis-embedding also increases users' uncertainty because it decreases the amount of information that is available about the vendor. Professionalism and reliability are examples of *trust-warranting properties* [Bacharach & Gambetta 2001] that might not be readily observable when interacting online. In an earlier study [Riegelsberger & Sasse 2001] we found that online shoppers feel it is more difficult to distinguish trustworthy vendors from untrustworthy ones on the Web than in a traditional shopping context. This may be partly due to a current lack of experience, but also because there are fewer cues available to base their expectations on. Furthermore, the cues that are available in an e-commerce interface are not necessarily firmly linked to trust-warranting properties. The size and state of the physical environment of a traditional vendor allows potential customers to assume that the vendor, is interested in future business. A vast e-commerce site, on the other hand, could just be a fabricated façade — the e-commerce equivalent of a *Potemkin Village*.

2.2 Affective Surface Cues

Apart from appearing competent (*ability*), trustworthiness is commonly communicated by showing that it is in the vendor's own long-term interest to behave as promised (*motivation* [cf. Deutsch 1958]). A good reputation for example signals good past performance and therefore professionalism, but it can also act as a sanctioning mechanism, as a vendor might not want to risk tarnishing it by disappointing or defrauding a customer. Trust based on such considerations is described as *cognitive trust* or *reliance* [Corritore et al. 2003; Lahno 2002] as it relies on assessing the incentive structure under which a vendor is operating. However, human trust-decisions are also based on affective reactions [Corritore et al. 2003; Einwiller 2001; Lahno 2002]. Research on consumer decision-making has established that purchase decisions are often based on immediate affective reactions that are rationalized *post-hoc* [Aaker 1996]. In his seminal paper '*Feeling and Thinking*', Zajonc [1980] posits that the most immediate reaction to any stimulus is along the dimension of *like* — *dislike*, and that the position on this continuum will influence the subsequent processing of the stimulus. Thus, it is not surprising that advertising these days largely relies on friendly faces and emotional imagery.

In the domain of e-commerce, the majority of studies focused on simple implementations of affective cues in the form of photographs. Fogg et al. [2001] found that photos can increase the credibility of online articles. Steinbrueck et al. [2002] found that an employee photo embedded on the homepage of an online-banking interface increased consumers' perception of trustworthiness. Zachar & Schaumburg [2002], on the other hand, could not find an effect on trustworthiness for an animated assistant. Trustworthiness and credibility in these studies were measured with questionnaires and rating scales. In a study based on qualitative interviews [Riegelsberger & Sasse 2002], we found a wide range of reactions, including very negative ones, to the different photos we tested. Our most recent study showed that photo effects depend on photo site interactions [Riegelsberger et al. 2003]. Table 1 gives an overview on the studies to date.

Authors	Description	Measurements	Results
Fogg et al. [2001]	Investigated how credibility of online articles varied in the presence of a photo.	Web survey with rating scales.	Photos increased credibility of articles.
Riegelsberger & Sasse [2001]	Embedded photos of editors, founder, customers and assistants in a mock-up of online retailer Amazon.de.	Focused Interviews, qualitative analysis.	Wide range of reactions, depending on photo implementation and user type.
Steinbrueck et al. [2002]	Compared mock-up of an online-bank with and without an employee photo.	E-Commerce trust questionnaire [Kammerer 2000].	Photo significantly improved trust perception.
Zachar & Schaumburg [2002]	Researched the effect of an animated assistant on trust in a mock-up of an e-commerce site.	Trust questionnaire by [Kammerer 2000].	No effect of animated assistant.
Riegelsberger et al. [2003]	Effect of 8 different photos of employees across 12 different vendor sites.	Decisions taken under financial risk.	Photo effects depend on site variables.

Table 1: Recent studies on the effect of affective cues on consumer trust in e-commerce.

Due to the great range in reactions to different implementations in these studies, we include two different photos in this study. These were selected based on results in a pre-study (see Section 3.1).

2.3 Effects on Visual Attention

Besides prompting immediate affective reactions, interpersonal cues such as photos of faces are also known to attract visual attention [Kroeber-Riel 1996]. While this effect is desired in advertising (where the advertiser has to compete with other stimuli for the recipient's attention), it might be counterproductive in the design of e-commerce systems. A user who cannot concentrate on his or her task due to too many affective stimuli being present might well decide to abandon a vendor's interface — not because of a lack of trust, but due to lack of usability. For trust-building surface cues to be successful in winning customers, it is important that they are perceived, but without interfering with the users' main task.

However, there is evidence that the findings from classic media do not directly translate to interactive media. The Poynter Project [Lewenstein et al. 2000] found that readers of online news largely focus on headlines and text and ignore photos and graphical information. Similarly, Benway's [1998] notion of *banner blindness* holds that Web users have learnt to ignore graphical and animated elements, as they consider them as non-functional advertisements. Pagendam & Schaumburg [2001] compared an 'aimless browsing' task and a search task. They found that recall and recognition for peripheral elements was higher in the browsing task. They take this as an indication that task type might mediate the effects of peripheral elements, and

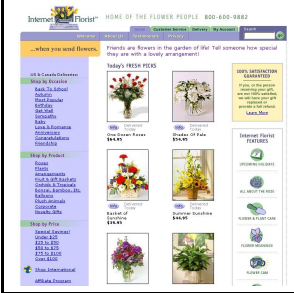


Vendor 1 (V1)	Vendor 2 (V2)	Vendor 3 (V3)
		
<p>'Good' Vendor Rating 10 (out of 10)</p>	<p>'Bad' Vendor Rating 4 (out of 10)</p>	<p>'Bad' Vendor Rating 4 (out of 10)</p>

Table 2: Vendors and their performance ratings (10 'good' / 1 'bad').

could also explain the differences in findings between classic and interactive media. A reader of a magazine might not be as task-focused as a user of a highly interactive online news site. While it is not clear that affective visual cues attract attention in e-commerce systems, many Web usability specialists recommend minimizing the use of non-functional elements [Nielsen 2000; Krug 2000]. This advice is supported by the very negative reactions of some users in our earlier study [Riegelsberger & Sasse 2001], who claimed that photos “clutter the page” and thus made it difficult to find the functional elements they were looking for. In this study, we investigate this claim by comparing versions of the vendors’ homepages, with and without photos embedded. Furthermore, with reference to Pagendarm & Schaumburg [2001], we compare two task types (‘get an impression of the page’ vs. search task; see Section 3.3).

3 The Study

3.1 Material

Flower delivery sites were used as material for this study. This domain poses a high non-monetary risk, such as embarrassment, if the service is not on time or the products are of lower than expected quality. Furthermore, quality cannot be assured by branded goods from trusted manufacturers, as is the case for buying electronics or computer hardware online.

For each vendor in the study, we had quality of service ratings that had been taken from the reputation sharing services Bizrate¹ and Epinions². These services rate online vendors based on post-order performance, handling of privacy and security, and customer satisfaction. Thus, these ratings are indicators for the presence of trust-warranting properties (see Section 2.1) that ensure shoppers against vendor-related risks that are present in e-commerce. As Bizrate and Epinions use different

¹<http://www.bizrate.com>

²<http://www.epinions.com>

#	Measure	Range	Description
M1	Preference	1 (bottom-ranked) 3 (top-ranked)	'Where would you feel most comfortable buying from?'
M2	Evaluation	1 ('bad') 10 ('good')	Rating of vendor on a scale from 1 ('bad') to 10 ('good') based on expected level of service and reliability.
M3	Estimate	1 ('bad') 10 ('good')	After a short explanation, participants estimated quality of service ratings as given by Bizrate and Epinions.
M4	Investment	0–100 Pence	Participants invested in vendors they expected to be 'good' vendors. Investments in a 'good' vendor earned them the amount they risked on top of their pay. Incorrect ones lead to a loss of the investment.

Table 3: Overview on measures used.

scales, we converted the ratings to a scale from 1 (poor rating) to 10 (best rating). On this scale, two of the sites were rated 4 and one site was rated 10 (see Table 2). As shorthand, we refer to the poorly rated vendors as 'bad' vendors and to the highly rated one as the 'good' vendor.

We mirrored the homepages and the first layer of the sites. This allowed us to vary the appearance of the homepage according to the experimental condition, while participants could at the same time explore the sites in depth. Information such as privacy, returns and shipping policy was accessible on all sites.

Two photos — one at a time — were embedded on the vendors' homepages. The photos had been selected based on the results of the study reported in Riegelsberger et al. [2003], in which 8 candidate stock photos were embedded on the three sites used in this study. We chose the photo that resulted in the highest ratings given to the sites, and the photo that resulted in the lowest ratings (averaging across the sites). The best-performing photo showed a young professional woman, the poorly performing one showed a man in his late fifties wearing a headset.

3.2 Measures

3.2.1 Preference & Risk-Taking

To date, research on trust in e-commerce has largely relied on qualitative interviews or questionnaires. However, eliciting trust responses through questionnaires and verbal accounts is limited by consumers' desire to appear as rational decision-makers. Hence, rather than asking people for several aspects of their trust reasoning — and possibly prompting factors that participants might not have considered in their everyday decision-making — we asked for simple decisions, some of them taken under financial risk. Table 3 gives an overview on the measures we used; below we introduce them in more detail.

For the preference measure (M1), we asked participants to rank the vendors they looked at in terms of shopping likelihood (assuming that prices are the same). This measure is of highest relevance for online vendors and it is also one users find easy to

Measure	Description
Scanpath	Total length of saccades (eye-movements) across the screen.
Backtracking	Backtracking occurs, if the angle between two subsequent saccades is acute ($<90^\circ$). It signifies a reversal in the direction of gaze.
Fixations in Region	Constant gaze of at least 100ms with a regional deviation of less than 7mm.
Binary Attraction	Whether a particular participant fixated a region of interest.

Table 4: Overview on eye-tracking measures taken.

make. To allow for a finer gradation of their responses, we also asked participants to rate the vendors' expected quality of service (M2) on a scale of 1 to 10. We used the simple anchors 'good' and 'bad'; as this reflects the way we think about simple, everyday decisions [Zajonc 1980]. Both measures were elicited twice: Initially after the first view of the homepage, and then again after the participants had explored the sites in more depth (see Section 3.3). This allowed for analysing the data for the stability of first impressions.

Glaeser et al. [2000] maintain that trust is best operationalized by having participants make decisions under risk. Such measures put more pressure on participants to distinguish 'good' vendors from 'bad' ones. We used two measures with financial incentives: Participants estimates of vendors' quality of service ratings (M3) and participants' willingness to risk parts of their participation pay on vendors (M4). These measures and their conceptual background are introduced in detail in Riegelsberger et al. [2003]; below we give a short overview: For measure M3, we explained to the participants that we had performance ratings for the vendors and asked them to guess the actual ratings on a scale from 1 to 10. To motivate the participants to estimate as accurately as possible, we added £3 to their base pay and deducted from this according to the squared error of their estimate. Measure M4 gave the participants the chance to invest up to £1 per vendor. Investment in the 'good' vendor was paid on top of their participation pay; investments in 'bad' vendors were deducted from their participation pay. As participants did not know the ratio of 'good' to 'bad' vendors, their investment could potentially impact their final pay in a range of £6. As our sample consisted mainly of students (for which these amounts have a high utility), this range created considerable risk. Measure M4 is based on the understanding of trust as 'willingness to be vulnerable': A participant who invests in a vendor exposes herself to risk for the chance to achieve a gain. We took these financial incentive measures only after the preference measures (M1 & M2) had been taken twice. By doing so we sought to avoid skewing the preference measures by introducing the financial incentives between the two measurements of preference.

3.2.2 Eye-Tracking

To identify potential effects of the photos on users' gaze pattern and on usability, we used the LC Eyegaze eye-tracking system. It allowed us to record a participant's location of gaze on the screen with a frequency of 50Hz. The LC Eyegaze system

#	Material	Task		Measurement	
T1	Screenshot of homepages	Just get an impression of the page. Decide which vendor you would be most comfortable buying from.	No Risk	Preference (M1) Evaluation (M2)	Eye Track
T2	Functional copies of vendors' sites	Now you can explore the sites in more depth. You have 2 minutes per vendor.	No Risk	Preference (M1) Evaluation (M2)	
			Risk	Estimate (M3) Investment (M4)	
(search tasks on other sites)					
T3	Screenshot of homepages	Find where you would go to order a bouquet of roses.			Eye Track

Table 5: Overview on experimental tasks.

is a remote tracking system that does not require the participant's head to stay in a fixed position. As a result, however, when participants change their body posture too quickly or move out of the tracking field, data is lost. We excluded participants whose gaze was tracked less than 90% of the time. We defined the area where the photo was embedded as a *region of interest*. The tracking software³ allowed us to calculate time spent looking at this region of interest, as well as the number of fixations and revisits to the region. Time spent in a region might not be a good measure to describe patterns of visual attention, as (a) there are great individual differences in processing speed and (b) due to differences in processing times of different types of visual content of a region. Thus, in this study we use a *binary attraction* measure. It states whether a particular participant fixated a region during a measurement period. Apart from region-specific measures, we also calculated the aggregate measures *scanpath-length* and number of *backtracks*. These measures are thought to be indicators for poor usability [Goldberg & Kotval 1998]. Table 4 gives an overview on all measures we calculated from the raw data stream.

3.2.3 Procedure and Tasks

When participants arrived at the lab, we briefed them about the study; then they completed questionnaires eliciting demographic data and information on their use of the Internet. The study consisted of three tasks (T1–T3, see Table 5). In Tasks T1 and T2 participants were asked to form an impression of the vendors. Both tasks were followed by eliciting their ratings and decisions (M1–M4, see Table 5).

In Task T1 the participants' exploration was limited to the vendors' homepages. In Task T2 they could explore the sites in more detail. Finally, Task T3 was a visual search task on the vendors' homepages. We recorded eye-tracking data for T1 and T3. As the eye-tracking system only allowed for displaying static pages, we could not record gaze data during the free browsing task (T2). For each

³EyeBrowse is Open Source software developed as part of the HIGHERVIEW research project. Please contact John D McCarthy (j.mccarthy@cs.ucl.ac.uk) for further information.

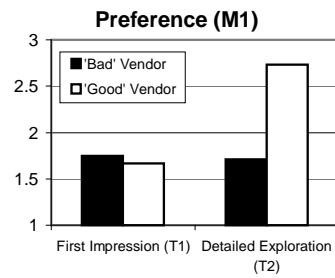


Figure 1: Preference measure (3 best, 1 worst). Difference significant for detailed exploration (T2; $U(23,16) = 52, p < 0.001$).

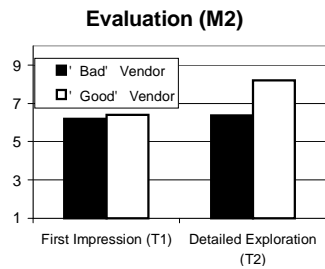


Figure 2: Evaluation measure (10 good, 1 bad). Difference significant for detailed exploration (T2; $t(37) = -3.6, p = 0.001$).

vendor's homepage, we created 3 versions in our experimental system: the original version, one with a 'trustworthy' photo, and one with a 'less trustworthy' photo (see Section 3.1). Each participant saw all three vendors, one with the trustworthy photo, one with the less trustworthy photo and one without any photo added. We had three different conditions, which — between them — gave us measurements for each vendor with each photo and no photo. Presentation order was counterbalanced.

4 Results

4.1 Participants

We had 39 participants (22 male, 17 female, average age 23 years). Most participants were students. 31 out of 39 spend two hours or more per day on the Internet and 30 of them had shopped online before. As such our participants are not representative of the current Internet user population, rather they represent the more experienced users we might expect in the future when consumers are collectively more Internet literate.

4.2 Telling 'Good' from 'Bad' Vendors

The first question of this study was whether users can tell 'good' from 'bad' vendors — where a 'good' vendor is understood to be one that is professional,

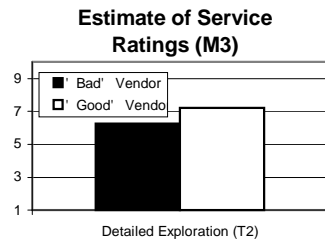


Figure 3: Difference for the financial incentive estimate measure (10 good, 1 bad; T2; $t(37) = -1.95$, $p = 0.06$).

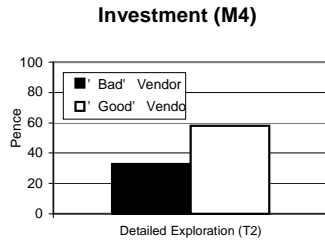


Figure 4: Significant difference in investment in vendor. (T2; $t(34.7) = 0.001$; using Levene's correction.)

reliable, and offers high level of privacy protection and security (see Section 3.1). Hence, we analysed the participants' responses in the baseline condition — without any photos added. We found that — purely based on a first impression (T1) — participants could not tell the poorly performing vendors from the 'good' ones (see Figures 1 & 2). However, based on a detailed exploration (T2), they rated the 'bad' vendors significantly worse than the 'good' vendor. This also holds for the investment measure (M4, see Figure 4), in which participants could risk parts of their participation pay. The difference is less pronounced for the estimate measure (M3, see Figure 3).

The difference between first impression and detailed evaluation can be illuminated further by looking at the number of participants who correctly ranked the 'good' vendor highest (see Table 6). Based on their first impression, only 5 out of 16 participants who saw the 'good' vendor in the baseline condition rated it highest. As there were 3 vendors in the study, this is no better than chance. After a detailed exploration, this number rose to 13, a result significantly better than chance ($p=0.0001$). This result also holds under financial risk (M4): 13 out of 16 participants risked parts of their participation pay with the 'good' vendor.

4.3 Effects of Photos on Trust and Preferences

On the 'good' vendor's site (V1), photos biased first impressions (T1), but there was no significant effect on the 'bad' vendors' sites (see Figure 5). However, this positive

	Incorrect	Correct
First Impression (T1)	11	5
Detailed evaluation (T2)	3	13

Table 6: Correct best ranks (M1) for the best vendor based on first impression vs. detailed evaluation (expected count > 5 for all cells, $\chi^2 = 8.13$, $p = 0.006$).

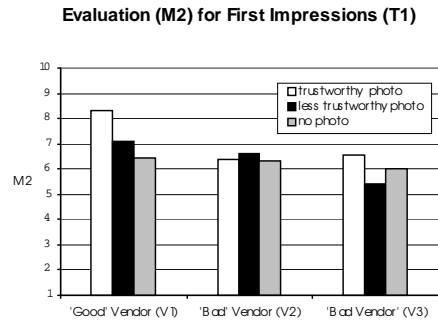


Figure 5: Effects of photo on first impression evaluation (Measure M2 1, bad – 10, good). Differences are significant for vendor 1 ($F(2,36) = 4.76$, $p = 0.015$).

bias on the ‘good’ vendor’s site was not strong enough to significantly increase the number of correct decisions based on first impressions. Analysing the data for the participants’ detailed evaluation (T2) we could not find any effect of the photos.

4.4 Eye-tracking and Task Performance

After removing participants with insufficient tracking accuracy, we had measurements for 25 participants for the impression task (T1). Participants did not significantly differ in the time they spent looking at the vendors’ homepages, nor did the presence of a photo have an impact on the time spent looking at the pages. Comparing the two photos does also not yield a significant difference: Participants looked as long at the ‘trustworthy’ photo as they did at the ‘less trustworthy’. The photos were fixated by the majority of participants: 21 out of 25 fixated the less trustworthy photo and 24 out of 25 fixated trustworthy one.

For the search task (T3) we had valid measurements for 29 participants. The presence or type of photo had no significant effect on the time it took participants to find their search target. Similarly, there was no significant difference in the time they spent looking at either the ‘trustworthy’ or ‘less trustworthy’ photo. However, comparing the binary attraction of the photos in the impression task (T1) and in the search task (T3) we found a difference (see Tables 7 & 8): While nearly all of the participants looked at both photos in the impression task, only one third looked at them in the search task.

‘Trustworthy’ Photo		
	Not fixated	Fixated
Impression task (T1)	16%	84%
Search task (T3)	62%	38%

Table 7: Percentage of participants who fixated the ‘trustworthy’ photo (expected count > 5 for all cells, $\chi^2 = 11.8$, $p = 0.001$).

‘Less Trustworthy’ Photo		
	Not fixated	Fixated
Impression task (T1)	4%	96%
Search task (T3)	76%	24%

Table 8: Percentage of participants who fixated the ‘less trustworthy’ photo (expected count > 5 for all cells, $\chi^2 = 28.4$, $p < 0.001$).

5 Discussion

5.1 Telling ‘Good’ from ‘Bad’ Vendors

We found that users can tell ‘good’ from ‘bad’ vendors based on an inspection of surface cues. This result is particularly convincing as we did not elicit users’ trust perceptions or attitudes, but their actual preferences. Furthermore, two trust measures (M3, M4) were taken under financial risk, lending additional credibility to the findings. Our results are encouraging for the development of trust in e-commerce: Based only on information given in the vendors’ interface, users could tell ‘good’ vendors from ‘bad’ ones. This suggests that surface cues are linked to trust-warranting properties, and that the surface cues are perceived and correctly decoded by users. Thus, as users collectively learn how to tell ‘good’ online vendors from ‘bad’ ones, general trust in the market place can be expected to grow: They see that they are capable of making correct trust-decisions and can thus iteratively risk more.

However, these encouraging results hold only for in-depth exploration (T2). Based on a quick glance of the home pages, our participants did not do better than chance when trying to tell ‘good’ from ‘bad’. There are several factors that may explain this finding: First, users — even our highly Internet-literate ones — have not accumulated as much experience with online shopping as they have with traditional shopping. Consequently, the processing of cues and subsequent decision-making are not as well trained and thus take longer. If this is the case, this problem is likely to diminish with time and experience. However, we might also face a persistent problem: The link between trust-warranting properties (such as professionalism and reliability) and cues may remain weaker on the Web than in the real world. The more important cues (i.e. those that are firmly linked to trust-warranting properties) may not lie in the spatial design of the homepage (the source of our participants’ first impressions), but in the interaction experience across the whole site; i.e. how the information is structured, how easy it is to find relevant policies, and how extensive the site is.

5.2 Effects of Photos on Trust and Preferences

We found that adding a photo to the homepages had an effect on participants' first impression (T1) of one out of the three vendors only. In that respect, our results support the earlier studies on affective cues in e-commerce by Steinbrueck et al. [2002]. As the positive effect was only observed on the 'good' vendor's site, the result opposes findings from our previous study [Riegelsberger et al. 2003], where — averaging across 12 sites and 8 photos — we found that photos had the most uniform positive impact on poorly performing sites. In that study, however, participants could instantly explore the site beyond the homepage and we did not split the measurement into first impression and in-depth evaluation.

The lack of effect for the in-depth exploration (T2) in the present study suggests that the effect of interpersonal cues depends on the depth of users' processing of these cues. This could also explain the contradicting results of previous studies (see Section 2.2). Affective cues might only influence users with low involvement [Petty & Cacioppo 1986] that don't go beyond the home page for their trust assessment.

The findings further support the notion introduced above that valid cues for trustworthiness are not to be found as much in home page layout and content but more in site structure and cognitive trust elements such as security and privacy policies. Users in this study were specifically asked to explore the sites in depth after their initial assessment and did better at telling 'good' from 'bad' vendors than users in the previous study [Riegelsberger et al. 2003]. Clearly, such a focus on in-depth exploration might not be found in field settings, where users might take decisions quickly.

5.3 Eye-Tracking and Task Performance

As in our previous eye-tracking study [Riegelsberger et al. 2002], in which we compared a page with a photo to one with a text box of the same size, we could not find any effects of photos on task performance in this study. Hence, photographs might not have an effect on task performance, but they might decrease user satisfaction, as users struggle to keep performance high.

An interesting result is the difference in the binary attraction measure between the impression task (T1) and the search task (T3). This effect could be due to repeated exposure. We previously found that participants in two successive goal-focused tasks ignored non-functional elements the second time they did a task on the same page [Riegelsberger et al. 2002]. In this study, however, several minutes passed between the two tasks, during which participants also completed search tasks on other pages. Thus, these results support Pagendam & Schaumburg's [2001] notion that a higher level of goal-direction in a task will lead users to be less susceptible to peripheral stimuli such as banner ads or photos.

6 Conclusions

6.1 Substantive Conclusions

With reference to the title of this paper, the clear answer is: *don't trust at first sight*. Users' first assessments of trustworthiness were not accurate and could be skewed by a simple intervention — the addition of a stock photo. Revisiting the aims of

this study in detail, our substantive conclusions are as follows: (1) Surface cues are linked to vendors' trust-warranting attributes and (2) users are able to perceive them; i.e. they are able to reach correct decisions about a vendor's trustworthiness based on cues present in the interface. However, (3) they are only able to do so after a detailed exploration of the interface.

With regard to photos as an example of affective surface cues that are supposed to build trust, we found an effect only on first impressions of the 'good' vendor's site (4). We could not find effects of photos on the correctness of users' trust-decisions, nor were the photo-effects consistent across different vendors. We could not find an effect of the presence of a photo on task performance. However, we found (5) that the photos attracted less attention in the search task than in the impression task.

Online vendors looking to increase consumer trust have to decide whether to invest in an impeccable e-commerce interface, or in other means such as classic advertising, or third party assurance mechanisms. Our results emphasize the significance of the interface as a communicator of trustworthiness. With regard to introducing affective trust cues, our recommendation is to be cautious — we only found effects based on first impressions, and those disappeared after participants could explore the sites in depth.

For researchers, the findings from this study encourage further investigation of surface cues, as they allow users to reach correct trust-decisions. The first question to address is which cues allow for a correct trust assessment. These are not necessarily those that are listed in designer's guidelines, as many of those cues might also be used by 'bad' vendors who want to appear trustworthy. On the contrary, it is desirable to find cues that cannot be easily included in a guideline on designing for trust. Identifying cues that cannot be easily mimicked could help to educate users about how to reach correct online trust-decisions, rather than only providing guidelines for vendors that are seeking to look trustworthy. Identifying cues that are hard to forge, and educating users about them, is likely to increase general trust in the whole market place.

6.2 Methodological Conclusions

In this study, we introduced several methodological innovations. First, we used existing vendors' sites whose level of service — based on reputation ratings — was known to us. This allowed us to compare participants' preferences and ratings to actual performance data. We could establish the correctness of participants' decision. A second advantage of this approach is that we could reward correct trust-decisions. We thus induced financial risk in our participants' decision-making, lending further credibility to the results (see Section 3.2). Finally, tests for trust were combined with a check for potential negative impacts on usability, the traditional core interest of the field of HCI. While this multidimensional approach surely can be further refined (see future work), we believe that it points in the right direction, as HCI has to combine an evaluation of classic usability criteria such as task performance with elements of user experience such as trust.

6.3 Future Work

The immediate questions that arise from this study need to be addressed: Trust based on first impressions should also be tested with a measure that induces risk (such as M4) to see whether the effects of photos hold for this stricter measure. Also, the discrepancy between user satisfaction (“photos distract”) and constant task performance needs to be further investigated, possibly using physiological measures to determine whether users have to work harder to keep task performance constant [Wilson & Sasse 2000]. Finally, the effect of task-type (impression vs. search task) needs to be tested in a study that counterbalances the order of the tasks. On a more general level, as said above, rather than identifying cues that can persuade users into trusting, research should be focused on identifying cues that enable users to reach correct trust-decisions.

Acknowledgements

This study was conducted as part of the HIGHERVIEW project, funded by BT. Jens Riegelsberger is funded by a BT studentship (WEB 164414/CT501045).

References

- Aaker, D. A. [1996], *Building Strong Brands*, Free Press.
- Bacharach, M. & Gambetta, D. [2001], Trust as Type Detection, in C. Castelfranchi & Y. Tan (eds.), *Trust and Deception in Virtual Societies*, Kluwer, pp.1–26.
- Benway, J. P. [1998], The Irony of Attention Grabbing on the World Wide Web, in *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, Human Factors and Ergonomics Society, pp.463–7.
- Brynjolfsson, E. & Smith, M. [2000], Frictionless Commerce? A Comparison of Internet and Conventional Retailers, *Management Science* **46**(4), 563–85.
- Corritore, C. L., Kracher, B. & Wiedenbeck, S. [2003], On-line Trust: Concepts, Evolving Themes, A Model, *International Journal of Human-Computer Studies* **58**(6), 737–758.
- Deutsch, M. [1958], Trust and Suspicion, *Journal of Conflict Resolution* **2**(4), 265–79.
- Egger, F. N. [2002], Consumer Trust in E-Commerce: From Psychology to Interaction Design, in J. E. J. Prins, P. M. A. Ribbers, H. C. A. van Tilborg, A. F. L. Veth & J. G. L. van der Wees (eds.), *Trust in Electronic Commerce: The Role of Trust from a Legal, an Organizational and a Technical Point of View*, Kluwer, pp.11–43.
- Einwiller, S. [2001], The Significance of Reputation and Brand for Creating Trust in the Different Stages of a Relationship between an Online Vendor and Its Customers, in M. Schoop & R. Walczuch (eds.), *Proceedings of the 8th Research Symposium on Emerging Electronic Markets (RSEEM2001)*. Available at <http://www-i5.informatik.rwth-aachen.de/conf/rseem2001/papers/einwiller.pdf> (last accessed 2003.05.26).
- Fogg, B. J., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., Boyd, J. & Brown, B. [2001], Web Credibility Research: A Method for Online Experiments and Early Study Results, in M. M. Tremaine (ed.), *CHI'01 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.295–6.

- Giddens, A. [1990], *The Consequences of Modernity*, Stanford University Press.
- Glaeser, E. L., Laibson, D., Scheinkman, J. A. & Soutter, C. L. [2000], Measuring Trust, *Quarterly Journal of Economics* **115**(3), 811–46.
- Goldberg, J. H. & Kotval, X. P. [1998], Eye Movement-based Evaluation of the Computer Interface, in S. K. Kumar (ed.), *Advances in Occupational Ergonomics and Safety*, IOS Press, pp.529–32.
- Kammerer, M. [2000], Die Bestimmung von Vertrauen in Internetangebote, Lizensiatsarbeit, Der Philosophischen Fakultät der Universität Zürich.
- Kroeber-Riel, W. [1996], *Bildkommunikation: Imagerysysteme für die Werbung*, Vahlen.
- Krug, S. [2000], *Don't Make Me Think: A Common Sense Approach to Web Usability*, New Rider Publishing.
- Lahno, B. [2002], Institutional Trust: A Less Demanding Form of Trust?, <http://www.uni-duisburg.de/FB1/PHILO/index/Lahno-Trust.htm> (last accessed 2003.05.26). *Revista Latinoamericana de Estudios Avanzados (RELEA)*. 16.
- Lewenstein, M., Edwards, G., Tatr, D. & DeVigal, A. [2000], The Stanford Poynter Project, <http://www.poynter.org/eyetrack2000/> (last accessed 2003.05.26).
- Luhmann, N. [1979], *Trust and Power*, John Wiley & Sons.
- Nielsen, J. [2000], *Designing Web Usability*, New Riders.
- Nielsen Norman Group [2002], E-commerce User Experience: Design Guidelines for Trust and Credibility, <http://www.nngroup.com/reports/ecommerce/trust.html> (last accessed 2003.05.26).
- Pagendarm, M. & Schaumburg, H. [2001], Why are Users Banner-blind? The Impact of Navigation Style on the Perception of Web Banners, *Journal of Digital Information* **2**(1). <http://jodi.ecs.soton.ac.uk/Articles/v02/i01/Pagendarm/> (last accessed 2003.05.26).
- Petty, R. E. & Cacioppo, J. T. [1986], *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, Springer-Verlag.
- Riegelsberger, J. & Sasse, M. A. [2001], Trustbuilders and Trustbusters: The Role of Trust Cues in Interfaces to E-commerce Applications, in B. Schmid, K. Stanoevska-Slabeva & V. Tschammer (eds.), *Proceedings of E-commerce, E-business, and E-government*, Kluwer, pp.17–30.
- Riegelsberger, J. & Sasse, M. A. [2002], Face It — Photos don't Make a Web Site Trustworthy, in L. Terveen & D. Wixon (eds.), *CHI'02 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.742–3.
- Riegelsberger, J., Sasse, M. A. & McCarthy, J. D. [2002], Eye-catcher or Blind Spot? The Effect of Photographs of Faces on E-commerce Sites, in J. L. Monteiro, P. M. C. Swatman & L. V. Tavares (eds.), *Proceedings of the 2nd IFIP Conference on E-commerce, E-business, E-government (i3e)*, Kluwer, pp.383–98.

Riegelsberger, J., Sasse, M. A. & McCarthy, J. D. [2003], Shiny Happy People Building Trust? Photos on E-commerce Web Sites and Consumer Trust, in G. Cockton & P. Korhonen (eds.), *Proceedings of CHI'03 Conference on Human Factors in Computing Systems*, *CHI Letters* 5(1), ACM Press, pp.121–8.

Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. [1998], Not So Different After All: A Cross-discipline View of Trust, *Academy of Management Review* 23(3), 393–404.

Shneiderman, B. [2000], Designing Trust into Online Experiences, *Communications of the ACM* 43(12), 57–9.

Steinbrueck, U., Schaumburg, H., Duda, S. & Krueger, T. [2002], A Picture Says More Than a Thousand Words —Photographs as Trust Builders in E-commerce Websites, in L. Terveen & D. Wixon (eds.), *CHI'02 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.748–9.

Studio Archetype & Cheskin Research [1999], eCommerce Trust Study, Studio Archetype has been taken over by Sapient. <http://www.studioarchetype.com/cheskin> where the report was accessible is no longer accessible.

Wilson, G. M. & Sasse, M. A. [2000], Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality, in S. McDonald, Y. Waern & G. Cockton (eds.), *People and Computers XIV (Proceedings of HCI'2000)*, Springer-Verlag, pp.327–39.

Zachar, T. & Schaumburg, H. [2002], Einfluss von Anthropomorphen Agenten auf das Vertrauen von Ecommerce Usern, in E. van der Meer, H. Hagendorf, R. Beyer, F. Krüger, A. Nuthmann & S. Schulz (eds.), *43. Kongress der Deutschen Gesellschaft für Psychologie*, Pabst Science, p.149.

Zajonc, R. B. [1980], Feeling and Thinking: Preferences Need No Inferences, *American Psychologist* 35(2), 151–75.