

Eye-Catcher or Blind Spot?

The Effect of Photographs of Faces on E-Commerce Sites

Abstract

E-Commerce is still hindered by consumers' lack of trust. In print advertising, trust and affective attitudes to brands are commonly built by using photographs of people. On the web photographs have been used only sparingly. Users as well as interface design experts believe that they decrease the usability of e-commerce sites as they attract attention and 'clutter' the interface. However, recent studies show, that they can be used to communicate trustworthiness. In this study we examined how participants' gaze behaviour when engaging in online shopping tasks is influenced by the presence of a photograph of a person. Users reported decreased task performance in the presence of photographs in previous studies. Our data does not support this claim. We found that the photograph attracts attention on first time view of a particular page structure. However, participants quickly learn the semantic structure of a page layout and ignore it on subsequent pages.

1 Introduction

In e-commerce, unlike in classic retailing, photographs have only been used sparingly. Human Computer Interaction (HCI) experts have argued for several years that photos and graphics should be minimised as they increase download times and act as distracters when users complete their tasks. Thus, they are said to decrease the usability of on-line services [15, 9]. However, bandwidth to potential clients is increasing, while 'lack of consumer trust' is still seen as an obstacle in the development of e-commerce [13]. Many studies have identified elements that can help to build trust in e-commerce. They are mainly cognitive cues for trustworthiness and professionalism. Examples include privacy statements, third-party certification and order tracking systems [4, 5]. However, trust comprises cognitive and affective dimensions [12]. Thus, we should examine how the affective impact of photographs, as it has been used in advertising for many years, can be used to build consumer trust on-line.

In this paper, we first introduce a conceptual basis for the use of photographs as trust-builders, drawing on expertise from advertising and sociology, and discuss results from studies on the effect of photos on perceived trustworthiness. We then present an eye-tracking study where we examined the impact of photographs of people on users' gaze behaviour and task performance. We will conclude with practical recommendations for designers of e-commerce sites and with suggestions for directions of further research on this topic.

2 Background

2.1 Photos in Advertising

When comparing print and poster adverts from the early days of consumerism to today's advertising, it is easy to see that the focus shifted from explaining product attributes to showing attractive people using the products. Why? The advertising industry has learned to make use of our impressive capabilities to process and store images, particularly images of people [8]. This capability is paired with a great weakness: we tend to interpret images, particularly photographs, in the same way we interpret the real thing. Whereas we critically ponder text we read, images we look at directly stimulate affective responses [14, 8].

What is true for pictorial information in general is even more the case for the processing of faces - be they real or on paper or on a screen. The neural area that deals with the processing of faces is distinct from the brain region that processes pictorial information [20]. These human capabilities and weaknesses have, often intuitively, been widely used in the advertising industry for many years.

2.2 Re-embedding: Trust in Distant Interaction

It is well known that interaction over distance requires more trust than face-to-face interaction [7, 11]. E-Commerce is no exception, as the persistent 'lack of consumer trust' on the Internet shows [13]. Trust in distant interaction is commonly supported through personal trust that is built through face-to-face interaction [7]. This process is called *re-embedding* [7]. Ordering books on-line and collecting them in the local bookshop, a service offered for instance by the German on-line-shop Libri, is an example of re-embedding in e-commerce. Based on the concepts of social presence [21] and telepresence [23], we have introduced the notion of *virtual re-embedding* [18]. We argue that even mediated human interaction can build trust that can be transferred to e-commerce systems. An example would be a personal customer service agent that can be contacted via chat or video-link from an e-commerce site.

2.3 Photos on the Web

As a first step to investigate the idea of virtual re-embedding, Riegelsberger & Sasse [19] as well as Steinbruck et al. [22] conducted studies to investigate the effect of photographs of people on an e-commerce site. Steinbrueck et al. found that a photograph significantly increased the perceived

trustworthiness of an on-line banking site. This result is corroborated by findings from Fogg et al. [6], who demonstrated that photos of authors can increase the credibility of articles in on-line magazines.

The findings of our qualitative study, looking at the effect of photos of people on the on-line shopping site Amazon.de are more controversial [19]. Reactions of participants ranged from welcoming their presence to expressions of increased mistrust, as they perceived the photographs as an attempt at trust manipulation. We categorised the participants based on their reactions in four distinct groups: *Relationship seeking*, *Function seeking*, *Benefit lacking*, *Mistrusting*. Relationship-seeking shoppers and benefit-lacking non-shoppers displayed overall positive reactions to the photographs. In the study reported in this paper we concentrate on the disapproval of the group we termed *function seekers*. Member of this group called the photographs a ‘nuisance’, as they ‘cluttered the screen without serving any purpose’. Their reaction supports the previously mentioned notion, expressed by many HCI experts, that photographs can decrease the usability of an on-line service. *Function seekers* claimed that finding functions and selecting products on an e-commerce web page is more difficult in the presence of photographs. Thus, we assume that an increased difficulty in completing a product selection or search task will result in an increased task completion time. Hence, our first hypothesis is:

H1: Task performance will be lower in the presence of a non task-related photograph than in the presence of non task-related text of the same size.

The participants’ claims are plausible, as many eye-tracking studies on print advertising have shown that photos, particularly faces (and here particularly eyes) get the first fixations when a page is viewed [8]. However, the Stanford Poynter project that researched on-line news reading does not confirm these results for on-line environments. On the contrary, their results show that users first focus on text, headlines in particular, and tend to ignore graphics and photographs [10]. The notion of *banner blindness*, introduced by Benway [2], further advocates caution when applying knowledge on eye movement from off-line reading to an on-line context. Benway found that users of web sites largely ignore ad banners that are designed to attract visual attention through imagery and animation [2]. However, these on-line studies have not specifically looked at photographs of people and faces, as they are being used in print advertising. Thus, based on the results of studies on print media and based on the claims of participants in our earlier qualitative study, we assume that photos of faces attract more visual attention than text.

H2: Users spend more time looking at a non task-related photograph of a face on an e-commerce page than on non task-related text of the same size.

The difference between the results of the Poynter study and the claims the participants in our on-line shopping study made may be explained by differences in tasks that were completed. In the Poynter study, participants were asked to read news rather than to complete an on-line shopping task. In a study in web banner ads, Pagendam & Schaumburg demonstrated that the type of the task that is performed by the user has a significant influence on recall and recognition of peripheral elements, such as web banners [16]. They compared a goal-directed search task to aimless browsing, where participants were asked to just have a look at the page to get a general impression. Based on studies by Benway [2] and Bachofer [1] they assume that goal-directed search relies on schemata that involve top-down information processing. Top-down information processing modes will result in less attention given to non task-related elements. Schaumburg & Pagendam assume that ‘aimless browsing’ is governed by bottom-up information processing, where visual attention is not guided by schemata but guided by external cues. Indeed, recall and recognition of peripheral banner ads was significantly higher in the ‘aimless browsing’ condition. Based on these results, we assume that the differences in information processing will also result in differences in gaze behaviour. Thus, we hypothesize:

H3: When users engage in a non goal-directed task, they will spend more time looking at photographs of faces than on text of the same size.

3 Method

3.1 Participants

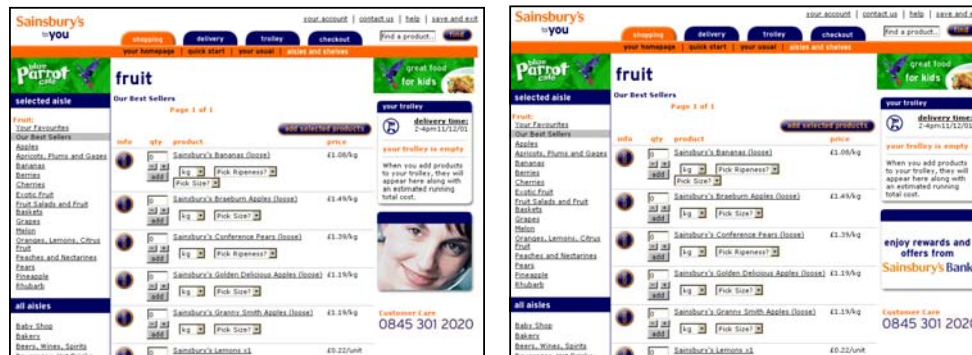
The study was conducted with 40 participants, students at the Department of Computer Science at University College London. They were paid £5 for their participation. The mean age of the participants was 22 years, 13 participants were female. The participants were very experienced Internet users, as 38 of them stated that they used the Internet for at least one hour per day.

3.2 Materials

We used pages from the on-line grocery-shopping site Sainsbury’s, a well-known British supermarket chain. The pages chosen were ‘aisle-pages’ containing product listing from different

product categories¹. Figure 1 shows an example of the pages used. Their structure was identical, just the products listed differed. For each page, we created a version with a photograph of a customer service agent (photo) and one with a box containing a text that detailed the services of Sainsbury's bank (text). Both photo and text were of the same size (160 x 135 pixels), subsequently we refer to this area on the screen as *region of interest*.

Figure 1: Example of 'aisle' page in photo and text condition



3.3 Procedure

When students arrived at the lab, they completed questionnaires on socio-demographic data and mood scales. Furthermore, we took baseline measurements of physiology data (Electrocardiogram (EKG) and Galvanic Skin Response (GSR)). We will report on results from these measurements separately. Then we calibrated the eye-tracker (LC Eyegaze). Instructions were presented to the participants on the screen. Once they had read and understood the instructions, they said 'ready' and we displayed the web page. On completion of their task they clicked the mouse. The web pages were incorporated as static screenshots, due to technological constraints of the eye-tracking software.

To compare the effects of different tasks we asked participants to perform 8 tasks on the on-line shopping pages. Table 1 gives an overview on the tasks.

¹ Task 7 was completed on a different page.

Table 1: Participants' tasks

TASK	
1. Select favourite product	5. Find 'help' section
2. Select favourite product	6. Find customer service telephone number
3. Find 'product search' function	7. Find recipe ideas ²
4. Find 'change delivery time' function	8. "Just look at the page" and get an impression

Tasks and conditions were counterbalanced. For every subject the conditions altered with every task. We duplicated the first task in order to allow us to make a meaningful comparison between first time view and subsequent views. For this task, also the stimulus page was counterbalanced. Table 2 gives an overview on the four different combinations of condition and task.

Table 2: Condition and task order for first and second view

		A	B	C	D
1. View	Task	Choose fruit	Choose fruit	Choose deli	Choose deli
	Condition	TEXT	PHOTO	TEXT	PHOTO
2. View	Task	Choose deli	Choose deli	Choose fruit	Choose fruit
	Condition	PHOTO	TEXT	PHOTO	TEXT

3.4 Measurements

The core dimensions that are evaluated in the field of Human Computer Interaction are task performance, user cost (comprising physiological and mental cost) and user satisfaction. Task performance has received most attention in the past. However, when we evaluate e-commerce systems that are targeted at potential shoppers, user satisfaction as well as user cost are at least equally important. The trust that users hold in an e-commerce site will impact user satisfaction, and can also minimise user cost. In the study we report here, we explored whether measures that aim to build trust, i.e. photographs of people, have adverse effects on task performance. The dependent variable measured for H1 is thus the time taken to complete the task. To explore the participants' gaze behaviour in more detail and to allow us to infer on effects on task performance and user cost, we also measured the absolute time of visual attention that was given to the region of interest. The LC Eye Gaze system allowed us to sample the position of the participants' gaze with a frequency of 50Hz. We report this measure as *sample count*, where each sample count equates 0.02 seconds of visual attention given to a

² This task was performed on a page with a different page layout. We introduced it to set the previous search-tasks that were conducted on the 'aisle-pages' apart from the non-search task, also conducted on the 'aisle-pages'.

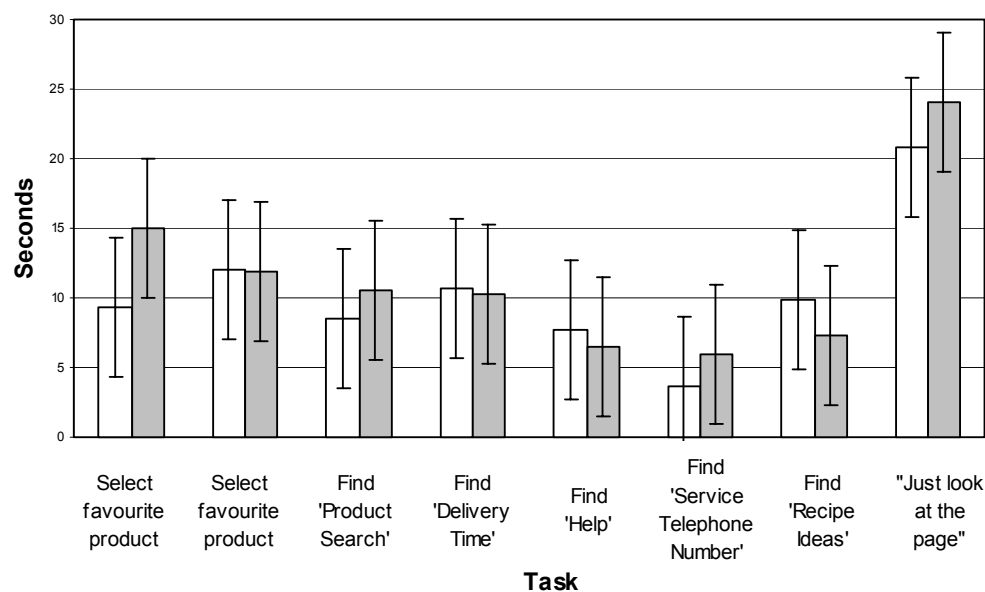
region of interest. We further analysed fixations: Fixations were detected by looking for sequences of gaze-point measurements that remain relatively constant. If a new gaze-point lies within a circular region around the running average of an on-going fixation, the fixation is extended to include the new gaze-point. The radius of the acceptance circle was set to 6.35 mm.

4 Results

4.1 Total Time (H1)

As figure 2 shows, there is no significant difference in the total time taken to complete a task between the photo and the text condition. There is thus no evidence that the presence of a photo compared to the presence of text has an influence on the time taken to complete a task. We cannot reject the null hypothesis, thus our results do not support H1.

Figure 2: Total time taken (white: text / grey: photo)

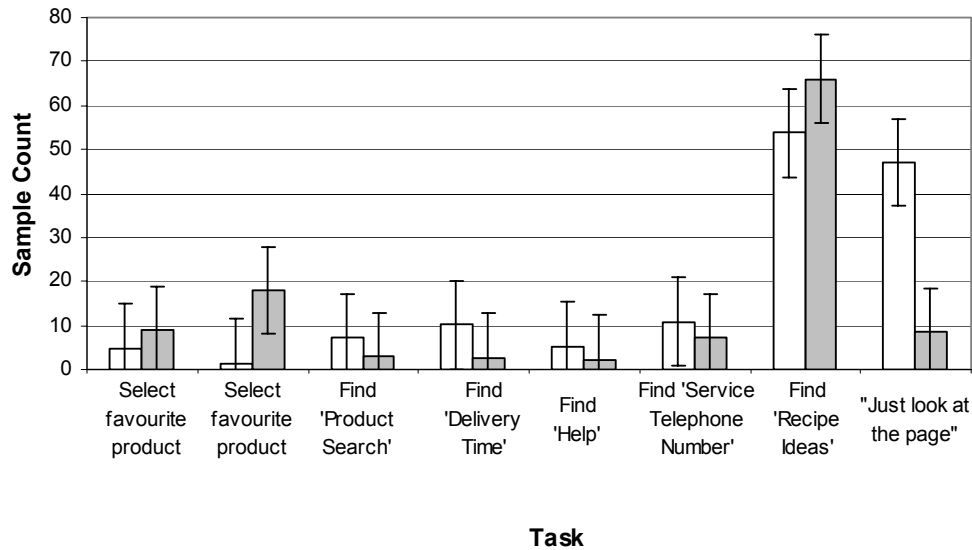


4.2 Time spent in Region (H2)

Again, over most tasks, there are no significant differences between the two conditions. Our data do thus not support H2. Participants do not significantly spend more time on photographs than on text. To the contrary, for most tasks, participants spend more time looking at the text than on the photograph.

Exceptions are task one and two, which both contain first time views of the pages (see section 4.4), and task 7 “Find recipe ideas”, which was conducted on a page with a different structure.

Figure 3: Time Spent in Region (white: text / grey: photo)



4.3 Impact of Task Type (H3)

Only for task 8 (‘Just have a look at the page’), the time spent looking at the region of interest differed significantly: Participants spent more time looking at the region in the text condition than in the photo condition. After excluding participants with a tracking accuracy³ of below 95% and removing outliers, we had 16 valid measurements for task 8⁴. Table 3 gives an overview on the results for this task.

Table 3: Results for task 8: “Just have a look at the page”

	Text	Photo	p	N
Sample Count	38.1	8.5	.015	16
Fixation Count	2.75	.75	.048	16
Relative Time	3.6 %	.5 %	.002	16

³ Tracking accuracy is number of registered samples divided by the total number of samples that could have been taken during task completion. Tracking accuracy is decreased by participants’ head movements or gazes away from the screen.

⁴ This task was not completed by all participants.

In the text condition, participants spent on average .76 sec on the region, whereas in the photo condition the average was .17 sec. The differences become even more evident when comparing relative times, i.e. the time spent looking at the region of interest relative to the time taken to complete the whole task. Thus, the data falsifies H3: For a non goal-directed task, participants spend more time looking at text rather than at photos.

4.4 Attraction of Region

One concern with above results is the extent to which time in a region is biased by the content of the region. As previously mentioned, it is known that face processing is fast and dealt with by specialised brain structures [20]. Thus using time in region might be an unfair comparison between the two conditions as it biased against H2 and H3. Thus, as a secondary measure we compared the conditions on the grounds of whether the region had been looked at or not. This measures the “attraction” of the region alone and is unbiased by differences in processing time.

This analysis revealed a difference between the first and second page participants viewed. When they look at the page the first time, significantly more participants look at the region when the photo is present (Pearson Chi-Square, $p=.027$). Looking at the second view of the pages, where participants again selected a favourite product, there are no significant differences, between conditions – the majority of participants ignore the region, be it text or photo (see table 4).

Table 4: Number of participants who viewed the region of interest

	TEXT	PHOTO	TOTAL
1st view			
Region NOT viewed	14	7	21
Region viewed	6	13	19
Pearson Chi Square: .027			
	TEXT	PHOTO	TOTAL
2nd view	15	16	31
Region NOT viewed	5	4	9
Region viewed			

5 Discussion

5.1 *Effects on Task Performance*

There is nothing in the data that allows us to conclude that the variance in time taken to complete a task is related to the presence of text or photos. Thus, the claims of participants in our qualitative study [19] are not corroborated. Similarly contradicting claims by users, Burke & Hornof, in a study on a visual search task, have found that animated banners do not decrease task performance [3]. However, they report that subjective workload, measured with the NASA Task Load Index (TLX), was increased. They conclude that in the presence of animations, task performance is held constant by users, albeit at the premium of increased mental cost. Based on our results and the claims of participants in our earlier study, we hypothesize that non task-related photographs have a similar effect.

The photos do not affect task performance, but they might affect mental cost, as users have to counteract this stimulus (bottom-up) driven guidance of attention to keep task performance high. Analysing the physiological measures that we took (Electrocardiogram (ECG) and Galvanic Skin Response (GSR)) will allow exploring the photos' effects on mental cost [25].

5.2 *Effects on Time in Region*

5.2.1 *Product Selection and Search Task*

When carrying out a product selection or search task, users do not spend significantly more time looking at non-related photos than at non-related text. Thus, our findings should raise concerns about directly transferring results from classic media (e.g. print) to on-line environments. Our results support the findings of the Poynter study [10] insofar as participants spend on average more time on the text than on the photo. We explain this result by the time it takes to discard a non-relevant object. As images are processed faster than text [8], users can discard non-relevant photos quicker than non-relevant text. Thus, less time is spent on photos, less visual attention is given to them. However, as indicated above, they might still increase user cost.

5.2.2 Non Goal Directed Task

We expected to activate schemata that are similar to off-line magazine reading by asking participants to 'just look at the site' without asking them to perform a search or product selection task. Thus, we assumed that for this task, the photo would attract more visual attention. However, participants spent significantly less time on the region in the photo condition than in the text condition. We explain this result by the fact that this task did not put time pressure on the participants, thus they could take the time to read the text, which they previously might just have skimmed.

On a methodological level, this result suggests we must clearly distinguish between different task types and processing modes when researching web site usage on such an atomic level.

5.3 Effect on Attraction of Region

We stated before that the measure of time spent at a region is biased towards text, as text takes longer to process, we have used the measure of *attraction*. We just measured whether a region had been looked at all. Again, across tasks, there were no significant differences between the attraction of photo and text. However, comparing the first time a page is viewed to subsequent views of the same page layout yields an interesting result: The first time they viewed the 'aisle-pages', the majority of participants looked at the region in the photo condition, whereas the majority did not look at the region in the text condition. Thus, for first-time viewing, results from off-line magazine reading translate to on-line shopping. Photos of faces do attract visual attention. They might thus be used as a modest means of virtual re-embedding.

A second interesting finding is that for subsequent views of the same page structure the region was largely ignored, independent of whether photo or text was displayed. We explain this effect through the participants' quick learning of the semiotics of the page structure. They learned where to expect task-related information. Unlike magazine-reading, where readers also look at photographs and adverts that are clearly identifiable by their position and are clearly not related to content they are reading, on-line shopper are more goal-focused.

6 Conclusions

6.1 Substantive Findings

This study is the first to look into the effect of photographs of people on on-line shopping sites. Evidence from related studies on the effect of photos on visual attention in other media is mixed

(section 2.1 and 2.3): Studies on classic media clearly state that faces attract visual attention. The best-known study for on-line media, the Poynter study, does not support these findings for on-line news reading.

Our most important result is that photos do attract more visual attention than text on a first time view of a page, when users perform a product-selection task. Steinbrueck et al. [22] found that one-time exposure to a photo of a person can already have a positive effect on attributed trustworthiness. Thus, experience from advertising can be used to give e-commerce sites affective attributes and ultimately build trust. However, our results suggest caution if a photo is used repeatedly on several pages of the shopping process. Firstly, it will largely be ignored and, secondly, ignoring it might come at a physiological cost to the users. They do perceive the pages as ‘cluttered’ while offering little functionality, as participants in our previous study stated. Furthermore, we found that on average, users spend more time looking at non-task related text than on non task-related photos (significant for task 8). This finding should also advocate caution when trying to communicate trustworthiness through more traditional and cognitive based approaches such as privacy statements and third party assurances that are usually implemented by using text. When trying to make it easier for users to perform their tasks on an e-commerce site, text should be as carefully placed as photos or graphics. Finally, our study showed that users are able to learn the semantic structure of a particular page surprisingly fast. This finding further emphasises the need for a consistent structure of pages across a site, as it has been advocated by many HCI experts [15, 9].

6.2 Methodological Findings

We argue that the time spent looking at an interface element, or the number of fixations on this element, is a problematic measure when inferring on its impact on the user’s information processing and – ultimately – task performance, user cost and user satisfaction. We compared photos and text, elements that are known to be processed in very different ways. Thus, we advocate the use of the percentage of users who looked at a particular region of interest, or *location of first fixation*, to measure the initial attraction of an interface element. Clearly, these have to be combined with physiological measurements to infer on user cost and with interviews, attitude scales or *Implicit Association Tests* (IAT) [17] to establish user satisfaction and – ultimately – perceived trustworthiness.

We have shown that results from on-line news-reading are not directly applicable to on-line shopping. Thus, results from eye-tracking studies from one domain of on-line services should be

generalised with great caution to other domains. Furthermore, even within the domain of e-commerce, our study showed that the task that is performed by the user on a particular page has an impact on gaze behaviour. Thus, for further studies it is advisable to differentiate the sub-tasks users perform on an e-commerce site and to evaluate them separately.

6.3 Future Work

Before we introduce thoughts for further research, we want to point out the limitations of this study. We opted for a controlled, experimental design to test the claims that have been made by users in our previous qualitative study. This meant that users could not explore the e-commerce site by selecting and clicking links as they would normally do. Thus, it can be argued that the study lacks ecological validity, a problem common to all tightly controlled experimental designs. However, unlike other studies, we used a non-invasive remote eye-tracking system, thus allowing participants to sit and move in a natural way. Hence, we believe that our results have some bearing for both researchers and practitioners at this early stage. They need however to be complemented by studies with less control and higher ecological validity that allow participants to navigate freely on an e-commerce site.

In order to allow for conclusions that are relevant for the design of e-commerce sites, the atomic measure of gaze behaviour should be combined with measurements of user cost and user satisfaction as well as perceived trustworthiness. In order to investigate the impact of photos on user cost, we will analyse the physiological measurements taken with this study. Perceptions of trustworthiness, attitudes and satisfactions can be elicited by using instruments from advertising and marketing research as well as qualitative methods from the social sciences (e.g. Grounded Theory [24]).

7 References

- [1] Bachofer, M. (1998) *Wie wirkt Werbung im Web?* Hamburg: Gruner & Jahr.
- [2] Benway, J. P. (1998) Banner Blindness: The irony of attention grabbing on the World Wide Web. 463-467. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.

- [3] Burke, M. and Hornof, A. J. (2001) The effect of animated banner advertisements on a visual search task. Technical Report. <ftp://ftp.cs.uoregon.edu/pub/hornof/CHI2002.pdf> .
- [4] Cheskin (1999) eCommerce Trust Study. <http://www.cheskin.com/think/studies/ecomtrust.html> .
- [5] Egger, F. N. (2001) Affective Design of E-Commerce User Interfaces: How to maximise perceived trustworthiness. Helander, M, Khalid, H. M., and Tham. 317-324. Proceedings of CAHD: Conference on Affective Human Factors Design. June 27-29, Singapore.
- [6] Fogg, B., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., Boyd, J., and Brown, B. (2001) Web Credibility Research: A Method for On-line Experiments and Early Study Results. 2001. CHI2001: Extended Abstracts. March, 31 – April, 5. Seattle.
- [7] Giddens, A. (1990) *The consequences of modernity* Stanford: Stanford University Press.
- [8] Kroeber-Riehl (1996) W., *Bildkommunikation* Munich: Vahlen.
- [9] Krug, S. (2000) *Don't make me think* Indianapolis: New Riders.
- [10] Lewenstein, M., Edwards, G., Tatar, D., and DeVigal, A. (2000) Stanford Poynter Project. <http://www.poynter.org/eyetrack2000/>.
- [11] Luhmann, N. (1979) *Trust and Power* Cichester: Wiley.
- [12] McAllister, D. J. (1995) "Affect- and Cognition-based Trust as Foundations for Interpersonal Cooperation in Organizations," *Academy of Management Journal*, vol. 38, no. 1, pp. 24-59.
- [13] Mitra, A. (2002) "Trust, Authenticity and Discursive Power in Cyberspace," *Communications of the ACM*, vol. 45, no. 3, pp. 27-29.
- [14] Nass, C. and Reeves, B. (2000) "Perceptual Bandwidth," *Communications of the ACM*, vol. 43, no. 3, pp. 65-70.
- [15] Nielsen, J. (2000) *Designing Web Usability* Indianapolis: New Riders.
- [16] Pagendarm, M. and Schaumburg, H. (2001) "Why are users banner-blind? The impact of navigation style on the perception of web banners," *Journal of Digital Information*, vol. 2, no. 1.

- [17] Plessner, H. and Barnse, R. (2002) Attitude Measurement Using the Implicit Association Test. *Experimental Psychology* 50[2].
- [18] Riegelsberger, J. and Sasse, M. A. (2001) Trustbuilders and trustbusters: The role of trust cues in interfaces to e-commerce applications. Schmid, B, Stanoevska-Slabeva, K., and Tschammer, V. 17-30. 2001. Norwell, Kluwer. Towards the E-Society: E-commerce, E-Business and E-Government. October 3-5, Zurich.
- [19] Riegelsberger, J. and Sasse, M. A. (in press) Face it: Photographs Don't Make Websites Trustworthy. 2002. CHI2002: Extended Abstracts. April 20-25, Minneapolis.
- [20] Rodman, H. R. (2002) Face Recognition. <http://cognet.mit.edu/MITECS/Entry/rodman>.
- [21] Short, J., Williams, E., and Christie, B. (1976) *The Social Psychology of Telecommunications* London: John Wiley & Sons.
- [22] Steinbrueck, U., Schaumburg, H., Duda, S., and Krueger, T. (in press) A Picture Says More Than A Thousand Words - Photographs As Trust Builders In E-Commerce Websites. CHI2002 Extended Abstracts. April 20-25, Minneapolis.
- [23] Steuer, J. (1992) "Defining Virtual Reality: Determining Telepresence," *Journal of Communications*, vol. 42, no. 4, pp. 73-93.
- [24] Strauss, A. and Corbin, J. (1998) *Basics of Qualitative Research* Thousand Oaks: Sage.
- [25] Wilson, G. and Sasse, M. A. (2000) Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality. Proceedings of HCI 2000: People and Computers XIV. McDonald, S., Waern, Y., and Cockton, G. September 5-8, Sunderland.