

Web Trustworthiness and Shyness

Web Trustworthiness and Shyness

by

Irvine F. Johnston

A thesis submitted to
the Faculty of Graduate and Post Doctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Canada

© 2012, Irvine F. Johnston



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-89341-8

Our file Notre référence

ISBN: 978-0-494-89341-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

This thesis explored the effects of colour and photos in Web design for the general population as well as for a particular target population, namely shy people. The thesis measured changes in the perceived trustworthiness and visual appeal of a Web page, for these two populations, produced by varying the colour of graphics, as well as the addition of photos of “friendly faces”, to home pages. The results showed that the colour of a simple home page made a significant difference in trustworthiness for a general sample of 84 undergraduate students. The pastel yellow page was rated significantly less trustworthy than any of the other nine pages. The blue page was rated more trustworthy than either the yellow or pastel purple ones. The results also confirmed that, for another general sample of 61 undergraduate students, some photos yielded higher trust ratings over the least trusted non-photo page. Conversely, the addition of some photos yielded lower trust ratings over the most trusted non-photo page. As well, the addition of a photo to an untrustworthy page was shown to increase trustworthiness, while the addition of a photo to a trustworthy page reduced trustworthiness. Trust scores were correlated with visual appeal scores, but the size of the correlation varied according to the various groups analyzed. Twenty non-shy participants rated the colours blue and pink significantly more trustworthy than twenty shy participants. Shy people were shown to tend to use the middle range of a semantic differential scale more than non-shy people. Taken together, the findings yielded differences in the predicted directions for the effect of colours and photos in Web design applicable to the general population as well as shy people.

Acknowledgements

A profound thank you to my thesis advisor, Dr. Gitte Lindgaard, for accepting me as a Ph.D. candidate, as well as your encouragement and wise advice, perceptive queries, insightful comments and suggestions. Also, thank you for the conversations; they were a very enjoyable and instructive part of these last several years. Your support led me to a fuller intellectual experience than I hoped for when I started this trek. Thank you to the members of my prospectus committee for your comments and suggestions. The thesis has grown, almost organically, but your original counsel always guided its path. Thank you to my external examiner, Dr. Elizabeth Sillence and my internal examiner Dr. Alejandro Ramirez. Thank you to the various professors who have introduced me to the application of social psychology methods to the field of Human-Computer Interaction. Thanks to Dr. Warren Thorngate and Dr. Avi Parush, who helped me to learn to write in different ways, a new experience for me. Thanks to Dr. Robert Biddle, who led me to see that early prototypes, and early failures, are not failings, but the essence of creative work. Thanks to the Doctoral Committee, Dr. Joanna Pozzulo and Dr. Robert Coplan, your varied interests have been the audience for whom I wrote this thesis. Thanks to my children, Greg and Tony, for their support and encouragement. Thanks to Dr. Lloyd Strickland for suggesting that a doctoral journey might be an enriching endeavour. Thanks especially to my wife, Chantal Leroux, ever supportive and encouraging, patient in reading my drafts for typos and poor storytelling, always reassuring when my confidence waned.

Table of Contents

Abstract.....	ii
Acknowledgement.....	iii
Table of Contents.....	iv
List of Tables.....	vii
List of Figures.....	viii
List of Appendices.....	ix
Introduction.....	1
A Trustworthiness Framework.....	3
Trust.....	5
Some Definitions of Trust.....	6
Stages of Trust Development.....	9
Distrust.....	13
Summary of the Concept of Trust.....	18
Measuring Trust.....	19
The McKnight Trust Scale.....	20
The Jian Trust Scale.....	22
Validity.....	26
What is Shyness?	29
The Measurement of Shyness.....	33
Sources of Shyness.....	34
Some Dimensions of A Model of Shyness.....	35
Concepts Related to Shyness.....	37
Introversion, Self-Esteem, and Social Phobia.....	38
Communication Apprehension.....	42
A Type of Attachment Bond.....	45
Summary of Shyness-Related Concepts.....	52
Shy People View Social Situations Differently.....	54
A Summary of Shyness.....	56
Trust, Visual Appeal, Colour, and Photos	57
Colour.....	57
Photos.....	64
Trust and Visual Appeal.....	67
The Visual Appeal Scale.....	68

Table of Contents (continued)

Statistics Used in this Thesis.....	69
Study 1.....	72
A Synopsis of the Pilot Study for Study 1.....	73
Formal Study 1.....	74
Method.....	74
Participants.....	74
Apparatus.....	75
Materials.....	75
Design.....	77
Procedure.....	77
Results.....	78
Discussion.....	82
Study 2.....	84
Method.....	85
Participants.....	85
Apparatus and Materials.....	85
Design.....	88
Procedure.....	89
Results.....	89
Data Exploration.....	90
Correlation of the Jian and McKnight scales Overall.....	90
The Jian Scale.....	91
The McKnight Scale.....	93
Correlations of Jian and McKnight Trust Scales.....	95
Correlations of 10 Items in Jian Scale with Item #11.....	96
Discussion.....	97
Study 3.....	99
Method.....	101
Participants.....	101
Sample Size.....	101
Apparatus and Materials.....	102
Design.....	104
Procedure.....	104
Results.....	105
Data Exploration.....	105
Comparisons of Pages based on Trust Scores.....	107
Comparisons of Pages based on Visual Appeal Scores.....	109
Correlations of Trust and Visual Appeal Scores Overall.....	111
Trust and Visual Appeal Comparisons for Shy and Non-shy.....	113

Table of Contents (continued)

Correlations of Trust and Visual Appeal for the Shy, Non-Shy, and Middle Groups.....	118
Summary of Results.....	120
Discussion.....	121
Study 4.....	123
Method.....	124
Participants.....	124
Apparatus and Materials.....	124
Design.....	125
Procedure.....	125
Results and Discussion.....	126
Study #5.....	128
Method.....	129
Participants.....	129
Sample Size.....	130
Apparatus and Materials.....	130
Design.....	131
Procedure.....	132
Results.....	132
Data Exploration.....	133
Comparisons of Pages based on the Less Trustworthy Pages.....	133
Comparisons of Pages based on the More Trustworthy Pages.....	134
The Effect of a Photo on Shy Participants.....	136
The Effect of a Photo on Shy and Non-Shy Participants.....	137
Correlations of Trust and Visual Appeal.....	138
Summary of Hypotheses.....	141
Discussion.....	141
General Discussion.....	143
Summary of the Main Findings and Contributions.....	143
Theoretical Implications.....	146
Future Research.....	152
Conclusions.....	155
References.....	157
Appendices.....	183

List of Tables

Table	Description	Page
1	Study 1. Friedman Ranks and Wilcoxon Results (row by column) for Trust scale.....	79
2	Study 1. Subjects preferring one page over another(Row Preferred over Column).....	80
3	Study 1. Correlations of Item 7, “corporate look and feel” and other Items for each Web Page.....	81
4	Study 2. Combined 22-item Scale.....	88
5	Study 2. Spearman Correlations of Jian and McKnight Scores by Web Page.....	91
6	Study 2. Friedman Ranks and Wilcoxon Results for Jian Scale.....	92
7	Study 2. Wilcoxon Results. Row Preferred over Column.....	92
8	Study 2. Friedman Ranks and Wilcoxon Results for McKnight Scale.....	94
9	Study 2. Friedman Rank Orders Studies 1 and 2.....	95
10	Study 2. Spearman Correlations of Rank orders Studies 1 and 2....	96
11	Study 2. Spearman Rank Correlation Coefficient of MCSD and Jian Trust Scores.....	96
12	Study 3. Five Hypotheses for Study 3.....	101
13	Study 3. RGB Definitions for the 10 colours in Study 3.....	102
14	Study 3. Jian Trust Scale t-tests.....	107
15	Study 3. Visual Appeal Scale t-tests.....	109
16	Study 3. Ranks of 10 Web Pages by Means for Trust and Visual Appeal.....	111
17	Study 3. Correlations of Trust and Visual appeal Scores by Colour.	112
18	Study 3. Pearson Correlations for Trust and Visual appeal for each Page within Shy, Non-Shy, and Middle Groups.....	119
19	Study 3. Hypothesis Number, Prediction, Test used, and Result.....	121
20	Study 4. Friedman Rank Order for the Three Photo Sets.....	126
21	Study 5, Five Hypotheses for Study 5.....	129
22	Study 5. Medians and Ranks for ArticleWorld, the Least-Trustworthy Page.....	133
23	Study 5. Median trust Scores and Ranks for Manitoba, the Most-Trustworthy Page.....	135
24	Study 5. Mann-Whitney Results.....	138
25	Study 5. Spearman Correlations of Trust and Visual Appeal for Each Page.....	139
26	Study 5. Hypothesis Number, Prediction, Test used, and Result.....	141

List of Figures

Figure	Description	Page
1	A Two-Dimensional Model of Trust and Distrust.....	13
2	Types of Validity.....	27
3	Attachment Styles.....	47
4	Attachment Styles based on Bartholomew (1990).....	50
5	Eight Web Pages for Pilot Study in rank order.....	76
6	Web Pages from Study 1 with mean ranks.....	78
7	Three Most Trusted Web Pages.....	80
8	Example Item in 22-item Trust Scale.....	87
9	Three Most Trusted Home Pages.....	93
10	Original Page #2 “Healthy Living” Home Page.....	102
11	Amended Healthy Living Page (Blue).....	103
12	Q-Q Plot for Page #6, Pastel Yellow, Jian Scores.....	106
13	Three Home Pages.....	108
14	Shy Trust Means for 10 Web Pages.....	116
15	Shy Visual Appeal Means for 10 Web Pages.....	117
16	Pastel Yellow Page.....	122
17	Six Photos used in Study 5.....	127
18	Low-and High-Trust Pages with and without Photos.....	131
19	Trust/Visual Appeal scores.....	140

List of Appendices

Appendix	Description	Page
A	Results of a Google Search, September 22, 2007.....	183
B	16-item scale for Pilot Study.....	184
C	Friedman, Kendall, and Wilcoxon Results for Pilot Study.....	186
D	Spearman Correlation of Rank Orders by Question Pilot Study..	187
E	Pilot Study for Study 1.....	188
F	Sample Recruitment Notice.....	197
G	Sample Informed Consent Form.....	198
H	Sample Opening Instruction Screen for participants.....	199
I	Sample Debriefing Form	200
J	8 Web Pages for Study 1.....	201
K	The Adapted 11-Item Scale for Study 1.....	203
L	Friedman, Kendall, and Wilcoxon Results for Study 1.....	204
M	The Adaptations of the Jian and McKnight Trust Scales	205
N	The Abridged Marlowe-Crowne Social Desirability Scale.....	209
O	Pearson Correlations of Jian Items with Item #11.....	210
P	Definitions for the Ten Colours in Study 3	211
Q	20-item Scale for Study 3.....	212
R	RCBS Shyness Scale.....	214
S	Study 3 t-tests, Jian Trust Scale.....	215
T	Study 3 Wilcoxon Results Jian Trust Scale.....	216
U	Study 3 t-tests, Visual Appeal Scale.....	217
V	Study 3 Wilcoxon Results, Visual Appeal Scale.....	218
W	Study 3 Minima, Means and Maxima for Shy and Non-Shy	219
X	Study 3 correlations for each page within Shy and Non-Shy groups.....	220
Y	Study 4. 27 Photos.....	221
Z	The Preparation of the “friendliness” Scale for Study 4.....	224
AA	The “friendliness” Scale Study 4.....	226
AB	Pearson Product Moment Correlations for items in the “friendliness” scale for Study 4.....	227
AC	Study 4 Friedman Ranks and Significant Wilcoxon Tests.....	228
AD	Photos for Study 5.....	229
AE	14 Web Pages for Study 5.....	230
AF	Study 5. Wilcoxon Results for Seven Pages.....	232
AG	Study 5. Results of Mann-Whitney Test of Shy versus Non-Shy Trust ranks.....	233
AH	Study 5. Median Scores for 14 Web Pages	234

Introduction

There is evidence that Web pages designed with certain colour combinations (Kim & Moon, 1998), or certain types of photographic images (Fogg, Marshall et al., 2001; Steinbruck, Schaumburg, Duda, & Kruger, 2002; Riegelsberger, Sasse, & McCarthy, 2003) may be perceived as more or less trustworthy than pages without these design features. Tens of millions of people are presently using personal computers to search out Internet-based information and advice for a variety of important and sensitive issues such as depression, anxiety, and stress (Pew, 2009). Shy people have been shown to use the Internet to establish online virtual friendships and to relieve feelings of loneliness (Ebeling-White, Frank, & Lester, 2007; Strizke Nguyen, & Durkin, 2004). Utz (2000) concluding an examination of online friendship found that the internet provided a venue in which shy people could find a safe place to develop relationships not available in the face-to-face world. Schank & Neaman (2001) see computer simulations in education as lessening the threat of public humiliation of students and thereby supporting shy people in learning. Web sites are available to help shy people with their struggles (Sites, 2011). Berger, Hohl, & Caspar (2009) found that Internet delivery of treatment for social anxiety disorder was statistically successful for 31 on-line subjects, of whom 85% were completely or moderately satisfied with the process. There exists a need to focus on specific guidelines that may assist web designers of chat rooms, and educational or therapeutic web sites targeted at shy people.

Beyond simple information, pages also exist for self-help to guide people in modifying behaviours such as alcoholism, anxiety, and shyness (Sites, 2011). Motivation to follow the advice of a change agent requires trust in that agent (Lewin, 1948). This trust will be a prerequisite for continued use of the page. The issue of credibility, especially in the formation of initial trust, is paramount in the use of these software products. This research explored the use of colour and photos in Web design for the general population as well as for a particular target population, namely shy people. Shy people are predisposed to evaluate faces and social situations differently from the general population (Battaglia et al. 2005; Pozzulo, Coplan, & Wilson, 2005). These predispositions may be disadvantageous to the development of trust in a Web page. This thesis measured changes in the perceived trustworthiness and visual appeal of a Web page produced by varying the colour of graphics, as well as the addition of photos of “friendly faces”.

Therefore, I chose to study the initial trust experiences of groups of shy and non-shy people accessing a “healthy-living” Web page. The variables of colour and photographic images were manipulated to alter these experiences.

The thesis comprised five experiments, identifying the effect of colour and photos on the perceived trustworthiness and visual appeal of a home page. The first established eight home pages with *prima facie* differences in perceived trustworthiness. The second validated a trust scale adapted from Jian, Bisantz and Drury (1998, 2000). The third investigated changes in the perceived “trustworthiness” and “visual appeal” of a Web page, for a general population as well as shy and non-shy people, when the colour of the graphics was manipulated. ,

The fourth established a set of photos of "friendly faces" as stimuli for the fifth, which investigated changes in the perceived "trustworthiness" and "visual appeal" for a general population as well as shy and non-shy people, of a Web page, with or without an attached photo.

I first present a framework for understanding trust, a framework which connects perceived trustworthiness with perceived expertise, to produce credibility. Second, I review the applicable research on trust and the stages of its development. Then, I discuss the relationship of trust to distrust, as this is an aspect of the existing research that is somewhat contentious. I finish that section with a discussion of the measurement of trust. Third, I examine shyness, its measurement, its sources, a model of its self-fulfilling dynamic, and a treatment of some related concepts. Fourth, I examine research supporting an expectation that shy people may perceive Web pages differently from the general population. Fifth, I review the research on the effect of colour and photos on perceived Web page trustworthiness and visual appeal. Finally, I present my five studies, a general discussion, and conclusions.

A Trustworthiness Framework

Fogg and Tseng (1999) propose that people assess a computer product along two credibility perspectives, perceived *trustworthiness* and perceived *expertise*. The combination of trustworthiness and expertise generates credibility, and this, following earlier social psychological experiments (Aronson, 1972; ch. 3), will influence our opinions and our actions.

Trustworthiness captures the perceived goodness of the particular aspect under investigation, while expertise captures the perceived competence of the

feature. A trustworthy person may be trusted, but may not have sufficient perceived expertise to be credible. On the other hand, a person can be an expert but be so untrustworthy that their word is doubted. This distinction is also true for computer hardware and software.

According to Fogg and Tseng (1999), the most salient aspect of the trustworthiness / psychological perspectives for the user is likely *on-screen characters*, the text, graphics, and images. This thesis studied the *trustworthiness* dimension as it is affected by changes in *on-screen characters*, specifically colour and the use of photos. Evidence that various colour combinations, as well as photos, may affect perceived trustworthiness, and a related feeling, resulting in a judgement of visual appeal is discussed later.

Importantly, perceived trustworthiness depends on the interpretation of prominent items in the interface (Fogg, 2003). Prominence describes the likelihood that a component on a Web site, such as text or a photo, will be noticed. Only when noticed can the item be interpreted and be assigned some trustworthiness rating. Culture, attitudes, and previously held views, in this theory, will cause different groups to interpret the same Web page elements differently. He uses the example of the various interpretations that can be taken, by different people, to the use of a Biblical text in a Web page. Biblical sayings, for some, are very trustworthy and possess expert knowledge, and are, therefore, very credible. Other people may not distrust the intentions of the Biblical writers, but may not assign any value to their expertise. Others may distrust the very intentions of the Bible, seeing it as authoritarian, therefore dismissing its very trustworthiness.

Fogg (2003) relied on this theory convincingly to explain how certain, apparently contradictory, research findings could be reconciled due to the variable level of prominence assigned by the investigators. Therefore, the need to notice something before it will be judged is obvious but may be a very powerful analytical tool. In one case (Princeton, 2002), people were telephoned and asked to evaluate what impact a privacy policy would have on perceived credibility. Sixty-five percent of respondents gave the presence of a privacy policy a high credibility rating. In the other case (Fogg, Soohoo et al., 2002) participants simply visited sites. Users were not told specifically to look for a privacy policy and few noticed it. Lacking perceived prominence, less than 1% of the respondents mentioned privacy policies. The vast difference between 65% and 1% may be explained by the fact that the users were not alerted to the privacy issue.

The importance of perceived prominence was seen to be central to this thesis' examination of the trustworthiness of Web pages as seen through the eyes of both shy and non-shy people. But, trustworthiness itself is a concept with a long history in psychological research. The next section reviews the most relevant aspects of that literature.

Trust

In this section, I examine some definitions of trust, as well as the stages of development of trust as it relates to Web pages specifically. As well, I discuss the relation between trust and distrust. Finally, I discuss some instruments to measure trust, finishing with one specifically focussed on trust in a non-human automated system, a machine.

Some Definitions of Trust

Trust has been an important variable in the social psychological study of mediated communication going back to Short, Williams and Christie (1976). Trust appears to have been largely an undefined term prior to the early work of Deutsch for the U.S. Office of Naval Research (as cited in Strickland, 1958). Later, Deutsch (1960) defined trust in terms of a scenario involving three components: i) a choice of action(s) that allows for both a desirable outcome and an undesirable outcome; ii) another person whose beneficial efforts are necessary for the desirable outcome to occur; and iii) a perceived “strength”, of the undesirable outcome that is greater than for the desirable outcome. Deutsch calls a set of actions in which the person chooses to act, hoping for the desirable outcome, a “trustful” act. Alternatively, a “distrustful choice” occurs if the person chooses to act expecting the undesirable outcome. It should be pointed out that the “strength” of the outcome is not the probability of the outcome. The perceived probability of the expected outcome is captured by the word “confidence” (Deutsch, 1960). Thus a trusting person has high confidence in a positive outcome, even in the face of a highly possible, very undesirable, outcome. Deutsch states that making the choice to take the ambiguous path is the behavioural exhibition of the personality characteristic. It would seem that an attitude of trust, which supports a trusting intention, is realized in trusting behaviour toward a trustworthy object. The decision to trust is supported by some calculus of pros and cons. Deutsch gives the example of trusting a baby-sitter: something bad could happen or something good could happen; the outcome is dependent on the baby-sitter; the bad outcome could be devastating, the good

outcome is simply the status quo. Although Deutsch identifies the path as “ambiguous”, it is the whole situation, the whole scenario, especially as regards the anticipated goal, the outcome, which is ambiguous. This definition is especially relevant for shy people, who have been shown to pay more attention to their expectations of negative effects in social situations relative to the general population (Asendorpf, 1987; Leary, Kowalski, & Campbell, 1988; Henderson & Zimbardo, 2001). Because of this, shy people may have a greater tendency to distrust.

The choice involved in the trusting action is supported by the trusting person’s attribution of a benevolent motivation on the part of the trustworthy other (Solomon, as cited in Strickland, 1958). Strickland (1958) expanded the range of supportive attributions in a study of supervisor surveillance and trust reports. He concluded that trusting behaviour could also come from an attribution of benevolent motivation directed not to the trusting person, but having, as a by-product, an expected positive outcome for the trusting person. He gives the example of a person trusting the Salvation Army to deliver clothes to the needy, not because the Salvation Army worker cares for the donor, but as an ancillary benevolence to their care for the less fortunate. Although the reasons may vary, the trustor expects positive actions from the trustworthy person, the trustee.

Rotter (1967), defined trust as “an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon”. At that time, he developed the Rotter Interpersonal Trust Scale, based on expressed trust for a wide range of “social objects” including

parents, teachers, physicians, politicians, classmates and friends. Although his definition of “trust” changed over the years, he continued to use this same scale for several research projects (Rotter, 1967, 1971, 1980). As an example, in 1980, he changed his definition to focus on the aspects of risk and ambiguity. His new definition for trust became “believing in others in the absence of clear-cut reasons to disbelieve” (Rotter, 1980). It would appear that “believing in” replaces the phrase “an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon”. Items in the scale address expectations: political, for example, “This country has a dark future unless we can attract better people into politics”; interpersonal, for example, “In dealing with strangers one is better off to be cautious until they have provided evidence that they are trustworthy”; societal, for example, “Hypocrisy is on the increase in our society”; and roles, for example, “Parents usually can be relied upon to keep their promises”. Rotter’s definitions, and his scale, include the positive expectancy component that Deutsch’s scenario-definition presented.

Rotter (1971) originally ignored the risk component in trust, but later (Rotter, 1980) added the property of “absence of clear-cut reasons to believe” to the definition. This is similar to Deutsch (1960) who remarked on the “ambiguity “of the possible outcome. The element of risk, for both researchers, is a necessary part of trust. Trust always involves the risk that the trusted person will not act in a beneficial way. If the trustee were guaranteed to be a benefactor, then there would be no need for trust. Interestingly, Deutsch (1960) commented on the “common confusion” over two applications of the word “risk”. When we trust, he said, we

risk a large possible loss against a small possible gain; when gambling, we risk a small loss compared to what we can win. The trusting person has a minimum of reason to choose to trust and a great deal to lose by trusting and yet, a high confidence in the trustworthy person.

Gambetta (2000) focussed on the *expectation* part of Rotter's (1967) definition of trust, the *confidence* noted in Deutsch (1958). Gambetta (2000) defined trust as "a level of subjective probability that we assign to the trustee". When we trust someone we implicitly mean that the probability is high enough that he will perform an action that is beneficial or at least not detrimental to us. Alternatively, when we say that someone is untrustworthy, the subjective probability is too low for us to trust. Thus, for Gambetta (2000) the concepts of trust and distrust involve probabilities, co-existing in the trusting person. Later, a more complete discussion of trust and distrust will be given.

All of these definitions decompose trust into these components: a choice; expected positive outcomes; the objective possibility of relatively high losses; and unclear reasons for optimism. The above definitions present a consensus on the aspects of the human experience of trust: a belief, an attitude, a personality characteristic, an intention, and finally, a behaviour. All but the last component was part of the shared framework connecting the reality to the words in the "Human-Machine Trust Scale" (Jian, et al., 1998, 2000), discussed in more detail later.

Stages of Trust Development

Jones and George (1998) discussed the day-to-day evolution of trust between two people, differentiating among three distinct forms of the trust experience:

distrust, conditional trust, and unconditional trust. They conceptualized trust as a dynamic experience in which we see a shifting from one state to another on the basis of changing experiences, values, attitudes, and emotions in the participants in a trust relationship. Their theory focussed on the ambiguity, the lack of clarity, the ambivalence in beliefs, attitudes, and expectations that often arise in complex and developing partnerships where roles are unclear and structure is uncertain. These are the same properties of trust, namely ambiguity and the high value of the potential loss, noted by Deutsch (1960). As well, this growth in trust can be conceptualized as a shift in subjective probabilities over time and events as proposed by Gambetta (2000). So, there is agreement that our experience is that person-to-person trust grows, and/or diminishes over time. This may apply as well to trust in a Web site.”

Sillence, Briggs, Fishwick, and Harris (2004) and Sillence, Briggs, Harris, and Fishwick (2007) studied fifteen women searching the Internet for information and advice about menopause. They explained some variation in the women’s Web-surfing behaviour by a three-stage trust development model. The women surfed the net freely during the first two one-hour sessions. They were then directed to specific Web sites for the last two one-hour sessions. The participants kept a log of their perceptions, and discussed their searches during 50-minute long meetings after each surfing session. As well, they verbalized their experiences as they surfed. The verbal protocols and discussions were transcribed and content analyzed. The number of times a particular criterion was mentioned was used as a measure of the importance of each criterion for the users. The researchers explained the women’s

behaviour by a three-stage progression toward trust acquisition. First, an affect-based choice was accomplished by rapid glances at a large number of sites. This first stage was primarily focussed on design (94%), and much less on content (8%). A later, second, stage saw time spent on a selected few sites, with attention here being primarily focussed on content (83%) as opposed to design (17%). Sillence et al. (2004, 2007) proposed that further site visits would cement the trust into a third, more solid experience, with continuing emphasis on the content of the site, especially its fit with the user's social identity. To that end, the women were given diaries and asked to return for follow-up interviews. The results for the second stage of trust acquisition were seen by Sillence et al. (2004, 2007) to involve a more cognitively intense, analytical process than the first.

In the second stage of trust development, the participants in Sillence et al. (2004, 2007) may have been using a slower process. This view was supported by Lindgaard, Dudek, Sen, Sumegi, and Noonan (2011), who asked 48 undergraduate students to rate 50 home pages on each of three dimensions: visual appeal, trustworthiness, and usability. They found that response times were longer for judgements of trust over visual appeal, indicating perhaps a qualitatively different process was in play. The first trusting stage seems to occur quickly, and is based on design characteristics such as “busy layout”, “too much text”, “boring web design”, “use of colour”. These descriptors are similar to certain characteristics found by Lindgaard, Fernandes, Dudek, and Brown (2006) to be highly correlated with visual appeal, namely interesting/boring, and good/bad colour. The research at hand, therefore, examined ratings of visual appeal as well as trustworthiness, to determine

the strength of the relationship between trust measures and visual appeal measures. According to Jones and George (1998) we achieve person-to-person trust in stages. The research by Sillence et al. (2004, 2007) suggests this is also true of trust in a Web site.

Fogg, Soohoo, Danielson, Marable, Stanford, and Tauber (2002) studied 2440 comments from 2684 people who were each asked to rank two Web sites for credibility and then make comments. Overall, 46.1% of comments used “Design Look” as the basis for initial credibility ratings. This work supported the findings of Sillence et al. (2004, 2007) who found the prominent aspects of a Web site at the initial stage of trust formation to be design, not content. However, there is a wide divergence with respect to the size of that aspect, 94% versus 46%.

Sillence, Briggs, Harris and Fishwick (2006) went on to validate this staged model of trust, naming the stages *heuristic analysis*, *content evaluation*, and *longer term engagement*. This study showed that the model could predict which of eight Web sites would be rejected in an initial 30-second viewing, as well as which would continue through the second stage to longer term engagement. The factors leading to trust or mistrust in the initial viewing of the site, such as layout, navigation, text size and quantity, advertising, and use of colour, identified by Sillence et al (2004, 2006, 2007), were similar to those identified in Fogg, Soohoo et al. (2002). One of the studies in this thesis used factors from both of these researchers to evaluate eight actual home pages specifically focussed on “healthy living”. That study specifically concerned the first stage in the trust acquisition process (Sillence et al., 2004, 2006, 2007).

Trust then involves a choice to do something based on an expectation of a good outcome with a real possibility that an alternative, harmful, outcome could ensue. Trust always involves some risk that the trusted person will not act in a beneficial way. That brings us into a discussion of distrust, which follows next.

Distrust

Some researchers have conceptualized the concepts of trust and distrust as subjectively assigned probabilities, co-existing in the trusting person, statistically envisaged along a single dimension (Jones & George, 1998; Gambetta, 2000). By contrast, other researchers have contended that trust and distrust are conceptually different (Robertson, Shaver, and Wrightsman, 1991; Lewicki, McAllister & Bies, 1998; McKnight & Chervany, 2001). Lewicki et al. (1998) viewing trust and distrust as separate constructs, present their view in four conceptual quadrants (Figure 1).

	Low Trust	High Trust
High Distrust	paranoia	trust but verify
Low Distrust	professional courtesy	interdependence

Figure 1. A Two-Dimensional Model of Trust and Distrust (Lewicki et al, 1998)

A one-factor model would easily explain the “paranoia” and “interdependence” cells. The other two, high trust/high distrust (“trust but verify”) and low-trust/low distrust (“professional courtesy”), go beyond a one-dimensional trust/distrust model. Lewicki et al. (1998) suggest that this two-factor model yields a richer understanding of the complex ambivalence found in the normal workplace.

They give an example of Boeing and a Japanese engineering firm, in which Boeing showed high trust, by working closely with the engineers, but, simultaneously showed high distrust, by not allowing them access to some secret areas in the buildings. Their view derives from inaccurate explanations for some valid research findings, and the definitions of trust and distrust used by the two-component theorists.

Lewicki et al.'s (1998) claims of experimental support for this model are not well supported. For example, they cite Constantinople (1969), who studied 952 undergraduate students over a four-year period, using a 60-item scale, five of which measured basic trust in self and five measured basic mistrust in self. An example of a trust item is "deep unshakeable faith in myself" which was measured on a 7-point scale from "uncharacteristic" to "characteristic". An example of a basic mistrust item is "pessimistic, little hope". Constantinople found evidence that basic trust in oneself grew over the four years while basic mistrust in self decreased.

Constantinople's research, based on an Eriksonian developmental psychology, which postulates trust/mistrust as a single dimension, showed a clear bipolar factor for "basic trust versus basic mistrust" (Constantinople, 1969, p. 360). How the Eriksonian infant *mistrust* relates to adult distrust is unclear, but, Lewicki et al (1998) equate the two. Constantinople (1969) lends support to a unidimensional bipolar model, thus discrediting the two-dimensional model. It seems to me that mistrust indicates a low level of distrust. To use the interpretation of Gambetta (2000), perhaps 90% trust would involve mistrust, whereas 60% trust might have distrust as its co-existing descriptor.

In a second example, Clark and Payne (1997) studied trust among a sample of 398 coal-mine workers. They found that some negatively-worded items, such as “In the future I will not expect to be rewarded for a job well done.” were negatively correlated with some positively-worded items, such as “In the future I expect to be able to rely on managers to be able to look after safety on the job.” They named the positive items “trust” and the negative items “distrust”. Using a multi-dimensional scaling analysis, they placed the various statements in a three-dimensional space citing two possible explanations for the negative correlations. First, they give examples from previous research (Dancer, 1985, 1986 in Clark & Payne, 1997) where a similar technical problem in facet analysis was dealt with by adding an explanatory variable termed item *format*. Facets are the constructs, proposed by an investigator, to explain the research findings (Guttman, 1954; Guttman & Greenbaum, 1998). The data in Clark and Payne (1997) can be explained by declaring “positive/negative format of item” as a separate dimension just as well as declaring “distrust” as a separate dimension. This renders the two-dimensional trust/distrust model unnecessary since this format facet replaces the second dimension of “distrust” posited by Lewicki et al (1998).

Another interpretation, offered by Clark and Payne (1997), is that some statements are focussed on *present trust* in managers and others are focussed on *future trust* in managers. The coal industry at that time was in a period of expected closures. It would be quite understandable if a statement such as “In the future I do not expect the pit to be successful unless management make better decisions.” would tap into a different dimension than statements such as “Management are

competent when it comes to matters of safety on the job.” The first is about management decisions and the real-world situation; the second is about trust in management decisions. As these are two qualitatively different judgements, it is reasonable to argue that a person’s trust in the success of their workplace should differ from their trust in management decisions in a specific, troubling time. Apparently, Clark and Payne (1997) do not support a two-factor model of trust/distrust. They do, however, lend support to the complex nature of trust, indicating that trust is volatile and likely to change, both in real-time and in different situations. They do not support the two-dimensional model proposed by Lewicki et al (1998).

A third study (Mancini, 1993) referenced by Lewicki et al. (1998) concerns a qualitative exploration of the trust relationship between journalists and politicians in Italy. Familiarity with politicians bears some risks since the reporter needs to be seen as independent, but also “in the know”. Both journalists and politicians know that the “rules” of familiarity and independence are broken frequently, so suspicion, “the very opposite of trust” (Mancini, 1993, p. 41) is an integral part of the Italian political journalism scene. Lewicki et al (1998) used this coexistence of trust and distrust, as an argument in favour of their two-factor model in which trust and distrust are perceived as distinct constructs. However, a unidimensional view allows for their coexistence as well. In fact, Mancini (1993) did say that there existed a “...continuing ambiguity between trust and suspicion...”. In a bipolar continuum such coexistence is quite understandable akin to a situation in which a

glass is one-third empty and two-thirds full. The co-existing probabilities of Gambetta (2000) is an adequate model for this coexisting ambiguity.

Finally, Lewicki et al. (1998) state that Robertson et al. (1991) separated trust and distrust along two dimensions. Robertson et al. (1991) statistically identified a set of positively worded statements and a set of negatively worded statements that factored out along non-parallel vectors. Lewicki et al. (1998) assimilated these sets under the rubrics *trust* and *distrust*. However, Robertson et al. (1991) simply pointed out that some statements, which happened to be negative, factored out differently from some positive statements. These factors could easily be labelled, “general cynicism” and “general hope”. These underlying dimensions of a person’s attitudes might alter trust in a particular situation, but they have not been shown to be trust or distrust. None of the research referenced by Lewicki et al. (1998) leads to a conclusion of trust and distrust being separable dimensions.

The definitions of trust and distrust given by Lewicki et al. (1998) are also a source of confusion in their search for some proof of a two-dimension theory. Lewicki et al. (1998) define trust as “confident *positive expectations regarding another’s conduct*”, while distrust is “confident *negative expectations regarding another’s conduct* (p.439).” Similarly, this definition of distrust involving a *confident* expectation of negative outcomes is not the co-existing “absence of trust” that is a property of our usual understanding of distrust. Given such a definition of distrust, it makes sense that two factors can be isolated. But in order to have applicability to daily human life, the words must be anchored in a common understanding. “How do people actually use these words?” is a necessary first

question to dealing with trust and distrust. Jian, Bisantz, and Drury (1998) asked people to rate 112 trust-related words as to whether they were perceived as “related to trust in general” or “related to distrust in general”. The results showed a high negative correlation ($r = -.96$) suggesting that trust and distrust are better understood as opposites, conceptualized along a single dimension, not different factors. This is how the concepts are used in this thesis.

Summary of the Concept of Trust

The view of trust taken for this thesis is a unidimensional one, with trust and distrust polar opposites, complementary probabilities representing a position along a single axis. Trust involves a risky choice, with the prospect of a desirable outcome and another highly undesirable, highly possible, outcome, both dependent on another person in whom we have no sure reason to believe. Trust is acquired in stages, both in person-to-person, and person-to-Web trust situations. Trust is context-dependent. The first stage of Web trust, primarily attending to design, is the focus of this thesis.

Trust always involves the risk that the trusted person will not act in a way beneficial to the trustor, and trust and distrust co-exist in the trusting person. Initial trust could be fragile if the perceived risk is high. This may be especially relevant for shy people, who pay more attention to their negative expectations in social situations relative to the general population. Because of this, shy people may have a greater tendency to distrust.

We will now examine a scale to measure trust, one based on a unidimensional model, seen to be quite adequate to describe trust and distrust.

Measuring Trust

There are many scales to measure trust, each one focusing on the particular viewpoint of the researcher (Wrightsman, 1991). As stated earlier, Rotter's (1967) Interpersonal Trust Scale focussed on the belief that another person's word or promise can be relied upon. It included items like: "Most elected public officials are really sincere in their campaign promises"; "Parents and teachers are likely to say what they believe themselves....". It measures trust with respect to people including parents, teachers, physicians, politicians, classmates and friends. Rotter's scale measures general trust towards others and society, but not towards specific people. In contrast, Rempel, Holmes, and Zanna (1985) have developed a scale for measuring trust in a specific person in a close relationship. An example of an item is "I can count on my partner to be concerned about my welfare." These scales were specifically focussed on human-human trust from particular perspectives. There exist, as well, trust scales designed to measure trust towards a Web site.

Several studies have categorized and sorted thousands of comments and evaluations in order to investigate the factors that would make a Web site more credible (Fogg, Soohoo et al., 2002; Fogg, Kameda, Boyd, Marshall, Sethi, Sockol, et al. (2002). The results of these large-scale online surveys were a very useful source of phenomenological insight into what factors influenced perceived trustworthiness across ten different categories of Web sites. These researchers, however, did not proceed to develop an instrument to measure trustworthiness. For the purposes of this thesis, these surveys were useful for guiding the measurement

in Study 1, which studied existing home pages with respect to prima facie trustworthiness.

The work of two groups of investigators who have designed instruments applicable to evaluating trust in the Web (McKnight, Choudhury, & Kacmar, 2002; Jian, Bisantz, & Drury, 1998, 2000) is most relevant to the goal of this thesis and is discussed next.

The McKnight Trust Scale

McKnight et al (2002) developed a 68-item scale to evaluate participants' trusting intentions toward e-commerce Web sites. The scale was based on their comprehensive model of organizational trust formation, including several high-level constructs. This model proposes that *Trusting Intentions* produce *Trust Related Behaviors*. *Trusting Intentions* are based in *Trusting Beliefs* which are fed by a general *Disposition to Trust*, which applies to individuals, and *Institution-Based Trust*, which is focussed on impersonal structures. In the case of e-commerce, this *Institution-Based Trust* is derived from Internet characteristics such as perceived safety and security. McKnight et al. (2002) demonstrated the convergent and discriminant validity of the 68-item instrument and the nomological validity of the model.

First-time Web site visitors, in that model, are in a stage of “initial trust” and McKnight, Cummings and Chervany (1998) posited that trust could form quickly, as has been found in “bricks and mortar” organizations.

This concept of initial trust aligns nicely with the model of initial trust proposed by Sillence et al. (2004, 2006, 2007), according to which initial trust

forms quickly, based on pre-existing dispositions. The participants in Sillence et al. (2004, 2006, 2007) chose to reject or accept Web sites based on some expectations with respect to good design.

The statements in the McKnight et al. (2002) scale were targeted specifically at a legal advice Web site, LegalAdvice.com, to investigate whether a consumer, investigating legal rights concerning a malfunctioning air-conditioner, intended to trust this site for a specific issue. The authors adapted existing human-human trust scales to capture those aspects that they judged most relevant to the Web situation. Among the many blocks in this model, two concepts, *Trusting Intentions* and *Trusting Beliefs*, were of particular interest to this study. *Trusting Intentions*, willingness to depend, uses phrases such as “would feel comfortable”, “always rely on”, and “could count on”. *Trusting Beliefs* is the extent to which the user believes this agency will be beneficial to them, and uses phrases such as “would act in my best interest”, “is truthful”, and “is sincere and genuine”.

The complete McKnight scale is based on the premise that the participants imagine themselves in a landlord-tenant dispute, and rate a legal advice Web site for credibility. Credibility involves both trustworthiness and perceived expertise (Fogg & Tseng, 1999). For that reason The McKnight scale was not suited to the goal of this thesis which aimed to measure initial trust in a Web page. However, portions of the McKnight scale were used in the second study of this thesis, which added some validation to the Jian Trust scale discussed next.

The Jian Trust Scale

Using an empirically based method, Jian et al (1998) began constructing their three-part Trust Scale by asking seven graduate students in English and Linguistics to describe trust in general, trust between people, and trust in machines. As well, the seven students rated 138 trust-related words collected from dictionaries and other studies as to whether they were negatively related to trust, not related to trust, or positively related to trust. Jian and his colleagues were searching for the implicit definition of trust held in the everyday use of the word. The written descriptions of trust generated 36 new words, but 74 were discarded, as they were rated as “not related to trust” by four or more of the seven raters, or were rated positively by some and negatively by others. Any words found in previous research, even if rejected by the process, were retained, resulting in a set of 112 words. They then asked 120 people to rate these words, as to whether they were perceived as “related to trust” or “related to distrust” with respect to “trust in general”, “trust between people”, and “trust in automated systems”. The trust/distrust results for each word showed a high negative correlation ($r = -.96$, $r = -.95$, $r = -.95$ respectively), suggesting that trust and distrust are better understood as opposites, not different factors. This, therefore, lends further support to the contention of this thesis that trust and distrust are best understood as two end-points on a single scale as argued earlier (pp.14 ff).

In further work, Jian et al. (2000) identified the words with the highest average ratings for “trust” and “distrust” across the three types of trust. Thirty words emerged, 15 weighing towards trust, 15 towards distrust. They then used a

paired-comparison analysis to rate the similarity of these words with respect to general trust, human-human trust, and human-machine trust. Based on a factor analysis and cluster analysis on the 435 pairwise comparisons made by 30 judges, Jian et al. (2000) were able to isolate nine significant factors underlying general trust, six for human-human trust, and eight for human-machine trust. These factors accounted for 77%, 77%, and 79% of the variance respectively. A cluster analysis grouped the words into 12 clusters according to similarity; 12 representative words were chosen, and an appropriate statement written for each.

This generated three trust scales, one specifically for trust for an automated system. This thesis aimed to measure a trusting attitude towards a Web page. Although the earliest discussions of trust (Deutsch, 1958, 1960) dealt with person-to-person trust, “trust” is also used with respect to other things, such as tools, animals, and institutions. These kinds of “trust” are often, though not exclusively, mediated by interpersonal trust, for example our trust in a bank is really trust in the bank personnel, trust in our automobile is trust in our mechanic. Alternatively, it is a derived use of the word trust. This was an issue for Jian et al (2000) who found no empirical evidence for equating human-human trust with human-machine trust. As it turned out, three of the five words people use most frequently to describe human-human trust are the same as those they use most to describe human-machine trust (Jian et al., 2000). These words are “trustworthy”, “loyalty”, and “reliability”. Interestingly, the words unique to automated systems were “honour” and “familiarity”; for people, they were “honesty” and “integrity”. So, the operationalization of trust-in-an-automated-system (Jian et al., 1998, 2000) is

related to trust in a person, but has some non-overlapping components. Because of the way the scale was developed, it has sound face validity. No empirical evidence is available for its reliability. However, evidence exists (Masters, 1974; Weng, 2004) that the number of items and response categories are sufficiently large to lead to satisfactory test-retest and acceptable Cronbach (1951) reliability values.

An interesting aspect of the genesis of the Jian scale involves the word “familiar”. In the first selection of words, some were retained, even though they had been rated as “not related to trust” by the seven graduate students. One of those words was “familiarity”. After the successive ratings of the remaining words as to their relation to trust, the word “familiar” was found to be significant, explaining 6.9% of the variance in the Human-Machine Trust scale. When cluster analysis was applied to the final 30 trust-related words 12 clusters emerged, but “familiar” was alone in a group. Galitz (1997) stated that if a screen is unfamiliar to you, or not easily assimilated, you use a slower conscious process to ensure that it is the one you want. Lacking this option, as in a quick, once-over impression of a page, the user may discard the site and label it untrustworthy. This decision to use the term “familiar” has received some support from an e-commerce study (Cheskin, 1999) in which participants evaluated the familiarity and trustworthiness of 102 Internet sites. The study found that the most trusted sites were for companies that were well known and most familiar. The least trusted companies were less well known (Cheskin, 1999).

Content validity comes from an evaluation of whether the scale attempted to represent the range of possible items concerning the measured reality. The

construction of the Jian scale (Jian et al., 2000) demonstrates that it explores the domain adequately, since:

- the initial words were carefully and expertly chosen;
- the number of these words was reduced i) by linguistic experts, and ii) by a large sample of educated people;
- further analysis with appropriate statistical tools identified 12 words, accounting for 79% of the variance.

This careful development lends a high level of assurance of content validity, that is, that the statements are representative of a universal pool of statements relevant to the concept (Cronbach & Meehl, 1955).

The content validity accruing to the original scale is based on expert ratings and subsequent factor and cluster analyses, of the *words* (deceptive, underhanded, suspicious, wary, harmful, confident, security, integrity, dependable, reliable, trust, and familiar) and *not the statements* later placed in the Human-Machine Trust Scale. The Jian scale was adapted for this thesis to evaluate initial trust in a Web page, but, it would seem that the existing content validation for the Trust in a Machine Scale would also apply to the trust statements in this adaptation. Two studies in this thesis examined the differences in initial trust for a Web page by shy and non-shy people. The *Human-Machine Trust Scale* (Jian et al., 2000) is well suited to this task.

The Jian scale was developed to measure general *Trust in a Machine*, but some question could still remain concerning its applicability to the present research. To address this issue, some added criterion validity was obtained by relating the

results on the Jian scale to those obtained from categories based in the work of Sillence et al. (2004, 2007) and Fogg, Soohoo et al. (2002), and from correlations with an instrument using some statements from the McKnight trust scale.

This thesis thus used an adaptation of the Jian scale to measure trust and included some statements from the McKnight scale in an attempt to further validate the adapted Jian scale.

Validity

Validity is an evaluation of the degree to which a measure actually represents the construct it is designed to measure (APA, 1966). The measure is used to draw conclusions about the concept of interest. For example, the measure might be a mark on a final examination in Grade 10 mathematics. The concept might be the expectation of success on the part of the student in Grade 11 mathematics.

However, some of the higher-performing students' performance will decrease, even to the point of failure; some of the lower-performing will improve, and some will receive about the same scores in the next grade. Thus, results from one year will often not predict next year's results perfectly. Validity is seldom 100%.

Therefore additional validation, especially as regards a particular context, is an ongoing process.

Validity allows the conclusions to be viewed by others as supported by the results and, therefore, dependable. Thus, some analysis of the validity of the scales was necessary to support a goal of my thesis, which was to use a trust scale to measure the perceived trustworthiness of various Web pages, modified by colour and photos. Validity is a single judgement having several, inter-related components.

(AERA, APA, & NCME, 1999). Originally, these were identified as content validity, construct validity, concurrent validity, and predictive validity (APA, 1954). In 1966, the APA combined concurrent and predictive validity under the rubric, criterion-related validity (APA, 1966). As experience increased, both criterion and content validity were seen as subtypes of construct validity (Anastasi, 1986; Shepard, 1993). An overview of the present view of the classification of validity is shown in Figure 2.

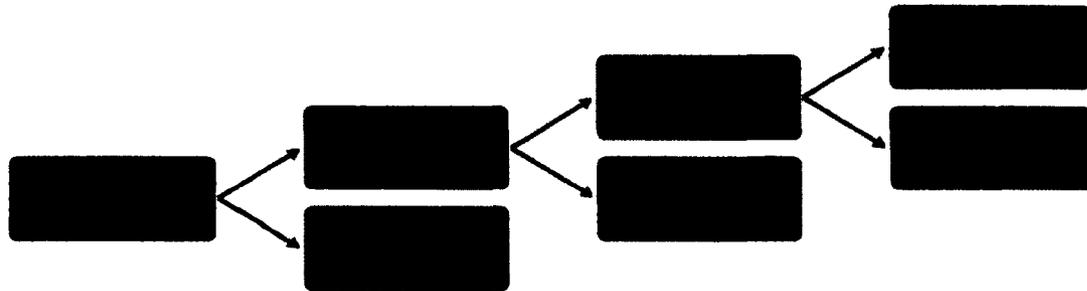


Figure 2. Types of Validity.

Construct validity assesses the level of correspondence between the measurement scale results and the concept being measured. Content validity is the degree to which the scale items are a representative sample of all possible items from a universal pool of items of interest (Cronbach, and Meehl, 1955).

Criterion validity is a measure of correspondence between the scale and other measures of the construct. Within this category, concurrent validity involves the use of a corroborating measure that measures the concept at hand. The correlation between measures can be positive, showing convergent validity, or negative, showing discriminant validity.

Predictive criterion validity is the ability of the measure to predict expected outcomes. It is in this area that the latest conceptualizations (Messick, 1995; AERA, APA, & NCME, 1999) have extended the concept of validity to include relevance for social consequences.

A common issue in assessing the validity of a scale is the fact that participants may be answering the questions based on some other underlying reality: e.g. they could be answering based on some desire to please the researcher, or give a socially acceptable answer. This response tendency would detract from the validity of the measuring instrument. A non-significant correlation with the Social Desirability Scale would indicate validity for the scale being evaluated, in that it is shown not to be simply a measure of a desire to answer in a socially desirable way.

The 13-item Shorter Marlowe-Crowne Social Desirability Scale (Reynolds, 1982) comprises a subset of statements (Appendix N) from an earlier 33-item scale designed to assess the respondent's desire to answer in a culturally appropriate and acceptable manner (Crowne and Marlowe, 1960). The Marlowe-Crowne scale was used extensively between 1960 and 1982 (Reynolds, 1982). Reynolds evaluated six short forms of the original scale to find a shorter version to increase its use to help validate research measures. Reynolds reported that the correlation between the 13-item short form and the full MC SDS was .93. The Kuder-Richardson Formula 20 figure for the 13-item short form was $\alpha = .76$, and $\alpha = 0.82$ for the full version ($\alpha = .88$ in Crowne and Marlowe, 1960). Reynolds concluded that his 13-item scale was superior to the other short forms based on reliability and convergent validity.

Zook and Sipps (1985) tested the 13-item scale on three groups and found the shorter version to possess average internal consistency ratings of .74. As well, they showed a 6-week test-retest reliability coefficient of .74. This 13-item scale was the one chosen for this thesis to guard against bias due to participants' desire to be viewed favorably by others.

What Is Shyness?

A significant part of the research in this thesis focussed on the differential response to stimuli by shy people compared to the general population. In this section, I therefore examine the causes and effects of shyness. I describe a model of its self-fulfilling dynamic and its measurement. Also, I examine what shyness is not, since there are many closely related constructs. To connect shyness to the aims of this thesis, I then review research examining the assumption that shy people may evaluate modified Web site home pages differently from the general population.

I came to be interested in shyness by a circuitous route. Originally, in response to a request for proposals from a Canadian Government agency to use social psychological methods to support middle-aged women re-entering the work force as entrepreneurs, I had identified shyness as a possible obstacle to this re-entry. I researched with the goal of developing web-based tools to help manage this shyness. This avenue of research proved to be an overly-ambitious task, but the lack of research on the initial evaluation of a web site by shy people became evident. I tightened my research focus to this area.

Shyness is a familiar, everyday word that has a variety of referents. First mentioned about 1000 years ago in an Anglo-Saxon poem, the adjective "shy" is

defined by The Oxford Dictionary (OED, 2012) as "easily frightened", "cautiously averse to encountering or having to do with some specified person or thing", and "shrinking from self-assertion". Henderson and Zimbardo (2001) state that, since 1977, the Stanford Shyness Clinic has defined chronic shyness as "a fear of negative evaluation that is sufficient to inhibit participation in desired activities and that significantly interferes with the pursuit of personal and professional goals". Shyness can be decomposed into *cognitive*, *affective*, and *behavioural* components (Crozier, 2005). The *cognitive* components include an augmented self-consciousness, in which shy people are more concerned with their own reactions than other people. As well, cognitively, shy people have more negative biases concerning the evaluations of situations and the attributions of causality. *Affective* variables involve items such as an amplified anxious feeling, and bodily reactions. Gaze-avoidance, and other reticent behaviours are subsumed under the *behavioural* rubric. Nearly all of us have experienced the sense of dread, racing heart, and attempts to become invisible that are examples of these three dimensions of shyness. These responses are very proactive when confronting an actually threatening situation. A certain amount of anxiety, with the accompanying physiological arousal, helps us to achieve some types of goals (Yerkes & Dodson, 1908; Zajonc, 1965). This is commonplace, for example, for athletes and students, politicians and teachers. The anxiety most of us feel towards public speaking causes us to prepare better, practice more, and thereby improve our performance. The shy person experiences anxiety more often, in more situations, and to a greater degree than the rest of the population. The anxiety is not resolved by practice. It stops shy

people from engaging in the feared activity. Shy people want to engage in social interaction, but are unable to, due to cognitive and affective components of shyness.

According to Heiser, Turner, and Beidel (2002) anxiety and shyness are related. Forty-one percent of a sample of 96 shy 20-year old college students also demonstrated social phobia or another anxiety disorder. Only 12.5% of 104 non-shy students did so. This anxiety can begin already in infancy (Kagan & Snidman, 1999). If a certain amount of anxiety, with the accompanying physiological arousal, is mixed with some positive affect such as interest, it will result in an ambivalent emotional experience, the transient *state shyness* with which we are all familiar (Asendorpf, 1989). This context-dependent shyness occurs in particular social situations, such as a junior prom, or meeting the Governor-General, and does not result in the person avoiding the event. However, transient shyness will modify the person's social behaviour. They do not make eye contact, and they avoid handshakes and small talk (Buss, 1986). As time passes, and the situation becomes more familiar, this passes, for small children as well as teenagers at a first dance, and adults in a new job situation. However, if anxiety is high across most, or all, social situations and has no accompanying positive affect, the stable syndrome of behaviours has been named *trait shyness* (Asendorpf, 1989). This continual shy behaviour can be a hindrance. Both trait and state shyness alters the person's attention and evaluation of social situations.

Buss (1986) distinguished two types of shyness, one based in fear, the other in extreme self-consciousness. The fearfully shy people begin their shy behaviour between six months and one year of age. This likely arises from a physiological

basis (Kagan & Snidman, 1999). Other children, at about six years of age, begin to worry too much about being evaluated, at the time when they begin to form a sense of self, (Buss, 1986). They fear the anticipated harsh evaluation. So, ultimately, in both groups, the shy people are fearful, some from a temperamental, others from a cognitive bias.

Zimbardo searched the academic literature in 1973 for research on adult shyness (Zimbardo, 1989). No substantial literature existed. As a result, Zimbardo and his students implemented a survey of open-ended questions, which later became a self-report checklist. This Stanford Shyness Survey (Zimbardo, 1989) has been administered to more than 5000 people in the United States and many other countries to assess the occurrence of shyness. Three-quarters said there had been a period in their life when they considered themselves to be a shy person. Only 6% said they had never experienced shyness. He then went on to ask the respondents to specify the components of this experience. Participants reported the physical symptoms of the “shy” experience to include silence, inability to make eye contact, blushing, heart pounding, and "butterflies in the stomach" (Zimbardo, 1989; Zimbardo & Radl, 1999). In order to gain more insight into the social difficulties of shyness, Leary (1986) proposed that the measurable behavioural components are anxiety and behavioural inhibition in social situations. Carducci (2002) identified the most obvious behavioural feature as silence, blushing, and stammering. Zimbardo (1989) stated that approximately 60% of the 5000 respondents, whose transient shyness was context dependent, identified these feelings as simply *their* reactions to *some situations*. However, based on these same, nearly universal

human experiences, about 40% labelled themselves as presently shy. Thus, it appears that the number of shy people is large.

The Measurement of Shyness

As discussed above, Zimbardo (1989) developed an in-depth instrument asking “are you shy”, “how shy are you”, and “what makes you shy”. It also asks about the occasions of shy behaviour as well as the physical, cognitive, and behavioural reactions. For this study, I was interested only in identifying a sample of shy people, not examining the wide range of issues in Zimbardo’s instrument.

Cheek and Buss (1981) developed a 9-item Shyness Scale that later was revised to form scales containing 11, 13, and 20 items (Cheek, 1983; Cheek & Melichor, 1985). The 13-item Revised Cheek and Buss Shyness Scale (Cheek, 1983), covers affective and behavioural components of shyness and focuses on being with people, asking for information, thinking of right things to say, and experience with the opposite sex. The RCBS (Appendix R) contains a rich series of statements that access the anxieties and inhibitions of shy people (Leary, 1991). This scale is widely used, and, according to Hopko, Stowell, Jones, Armento, and Cheek, 2005, has acceptable internal, and test-retest reliability. Leary assessed the RCBS scale, amongst several other measures focussed on different aspects of social anxiety, and found the RCBS to be “the measure of choice” when both subjective anxiety and behavioural inhibition components of shyness are of interest. This aspect of construct validity is its content validity.

Leary (1991) gives evidence of convergent validity between the RCBS and several other scales: the Social Avoidance and Distress scale (Watson & Friend,

1969), the Interaction Anxiousness scale (Leary, 1983a), the Social Reticence Scale (Jones & Briggs, 1986); as well as qualitative ratings by close friends and family. He gives evidence of discriminant validity with “fearfulness” scale, self-esteem (Jones, Briggs, & Smith, 1986), and extraversion (Eysenck & Eysenck, 1968).

Sources of Shyness

Zimbardo (1989) reported that 73% of 2482 American students who were asked if they had ever been shy stated they had been shy *at some point in their lives*. Sixty-eight percent of 540 U.S. Navy personnel (mean age 26) gave the same response. Forty-two percent of the students, and 33% of the Navy personnel reported that they were *presently shy*. Very few said they had never experienced shyness: 7% and 9% respectively. Some of this is likely biologically rooted (Kagan & Snidman, 1999; Buss, 1986), and some is learned (Buss, 1986). Cross-cultural studies (Zimbardo, 1989; Carducci & Zimbardo, 1995; Zimbardo & Radl, 1999) show that some is based on cultural norms.

With respect to biological roots for shyness, about 20% of infants have been found to display a heightened sensitivity to sounds and smells (Kagan & Snidman, 1999). This temperamental bias predisposes them to be fearful of unfamiliar situations in later life. However, for the remaining 20-30% who report shyness, the anxious feelings would seem to be acquired through learning.

Cross-cultural studies shed some light on these mechanisms (Zimbardo, 1989). In Japan, 60% of the 305 respondents reported being presently shy, while in Israel only 31% of 231 did so. For 75% of the Japanese respondents, shyness was a personal problem. It was a problem for only 42% of the Israeli respondents.

Zimbardo theorized that this might be explained by societal rearing practices. Any success by a Japanese child is attributed to the actions of parents, grandparents, or teachers. Any failure, on the other hand, is the student's fault. As a result, when presented with a potential interaction, Japanese children have learned to expect high losses for failures, and no gains for successes. This could cause anxiety for the Japanese child. Israeli children, conversely, are rewarded just for trying (Carducci & Zimbardo, 1995; Zimbardo & Radl, 1999).

Findings such as these have been used to develop a model for the development and maintenance of the various components of shyness. It was from the dynamics of this model, discussed next, that some of the hypotheses for this thesis arose.

Some Dimensions of a Model of Shyness

The Stanford Shyness Clinic has developed a model of shyness based on 17 years' experience (Henderson & Zimbardo, 2001). This model identifies three dimensions to shyness: the characteristics of the individual, the demands of the social environment, and the goals of the individual. This perspective drives the various techniques used to modify the appropriate components of shyness. Researchers of social phobia (Turk, Lerner, Heimberg, & Rapee, 2001; Otto & Safren, 2001) and shyness (Henderson & Zimbardo, 2001) have further specified the dynamic of the various factors along these three dimensions. Turk et al. (2001), although reporting research with people who are socially phobic, propose that their model can apply to shy people as well. Although the exact relation between social

phobia and shyness is questionable (Beidel & Turner, 1998), certain features are common to both and will assist in understanding shyness.

Research shows that shy people perceive things differently from the general population (Turk et al., 2001; Otto & Safren, 2001; Henderson & Zimbardo, 2001). First, socially anxious people expect to be seen as inept, and fear this expected disclosure. Zimbardo (1989), comparing the shy situation to that portrayed in the classic Stanford prisoner/guard experiment (Haney, Banks, & Zimbardo, 1973), theorized that shy people are inhibited by a "guard-self" that forecasts that they will look ridiculous, others will laugh, or that the place or time is inappropriate. This inner process promises safety only if the shy person remains unseen, unheard.

Second, this anticipated humiliation causes increased anxiety that seems overwhelmingly to support attention for negative signals. In any specific, real human interaction, the shy person experiences real anxiety due to these real-time perceptions of humiliation.

Finally, the shy person, to curtail this anxiety, does not engage in any mitigating strategies, but completely ends the interaction. However, this act is seen as a failure, and supports the original expectation of interactive ineptitude. The short-term acute anxiety is avoided, but the chronic expectancy of failure is strengthened, ready for the next real human encounter. One may speculate that this cycle could lead shy people to notice items on a Web site differently from the general population. It could cause them to choose a Web site in a much more pre-emptory way than the general population.

Some points need to be made to flesh out this model. First, the actual level of ineptitude of the shy person, or of derision by the audience, is not the source of shyness. The subjective evaluation of the shy person her/himself is the engine that drives the initial expectation and the person's evaluation of the failure in the real encounter. Second, even when shy or socially phobic people use skills to diminish anxiety, they still see this as a failure and the improved skill set can become a confirmation of their ineptitude (Phillips & Metzger, 1973; Wells et al., 1995). This can also be true for anxiety-easing drugs. Third, people in general may have a predisposition to notice negative more than positive social cues. This may have a survival advantage (Rozin, 1986; Zahorik, 1979). Shy people may be even more attuned in this way than others either through biological or environmental factors (Buss, 1986).

This thesis used a shyness scale that measured the affective and behavioural components of shyness. Equipped with this model of shyness, we will next examine some psychological constructs that are related to, and sometimes confused with, shyness. The aim of this section is that, by looking at these parallel or overlapping constructs, we will more clearly understand shyness.

Concepts Related to Shyness

Leary (1983b) cautioned that we should distinguish the various shyness-related terms (e.g. reticence, shyness, communication apprehension, speech anxiety) carefully, since any confusion will result in measurement problems. For that reason, it is necessary to discuss concepts that are closely related to shyness, to ensure that the distinctions between these are clear to the reader.

The concepts of introversion, low self-esteem, and social phobia are related to shyness. As well, researchers have identified a category of behaviour termed “communication apprehension” that is said to be a “label” for shyness (Rapp, 1999). As well, the *theory of attachment* that explains adult social interaction strategies based on experiences in infancy (Bowlby, 1982; Bartholomew, 1990) is discussed here. The following sections discuss these five constructs: introversion, low self-esteem, social phobia, communication apprehension, as well as infant-caregiver attachment, and their relation to shyness.

Introversion, Low Self-esteem and Social Phobia

Introversion is different from shyness, although, statistically, they are moderately related (Heiser, Turner, & Beidel, 2003). It is the nature of introverts to turn inwards towards their inner experiences (Jung, 1923). Extroverts, at the other end of this continuum, focus on the outer objects. Many people lie in between, responding to both inner and outer experiences. Introverts can interact with other people without the fear, self-consciousness, and negative thoughts that shy people experience. As well, introverts choose to work alone, to pass their leisure time alone. The introverted person prefers solitary activities, while the shy person fears social activities. Although the majority of shy people are also introverted, there are shy extroverts who are privately shy and publicly outgoing (Zimbardo & Radl, 1999).

In this thesis, I was interested in examining the reactions of ‘people who perceive themselves to be shy’ to various Web page formats. Hardin et al. (2006) used the Revised Cheek and Buss Shyness Scale (Cheek, 1983) and the Eysenck

Personality Questionnaire (Eysenck & Eysenck, 1991) and found a significant negative correlation ($r = -.43, p < .001$) between extraversion and shyness. I expected that some shy participants in my experiments would be introverts, but the fact that they are shy, and therefore might be interested in accessing a shy self-help, Web site was of interest. For that reason, I did not assess introversion.

Self-esteem measurements attempt to ascertain an individual's self-evaluation of their own worth (Rosenberg, 1965). In a study by Crozier (1995), shyness showed a significant negative correlation ($r(135) = -.62, p < .001$) with global self-esteem for 137 nine-to-twelve year olds. To further explain the relation, Schmidt and Robinson (1992) found that fearfully shy people self-reported lower self-esteem on the Coopersmith Self-Esteem Inventory (1981) than those who were self-consciously shy ($p < 0.01$). Shyness and self-esteem are negatively related overall, but, within the total group of shy people, there exists significant variation between the two shy types.

Moreover, Carducci (2002) proposed that shyness involves the whole self, while self-esteem can vary depending on the aspect of the self in question. Shyness certainly involves the whole self as it engages in any social interaction. To that extent, and since social interaction is a large part of human daily life, shyness and self-esteem will be negatively correlated. As we isolate certain of the activities of shy people, such as solitary activities or ones where functional coping strategies are used, this correlation will diminish or even vanish. In support of this, Prakash and Coplan (2003) found no overall correlation between shyness and *athletic self-esteem* in a group of adolescent competitive female skaters. However, skaters who

coped with stress by focussing on their emotions had a negative correlation between shyness and athletic self-esteem. For those who coped by avoiding the stressor, or using problem-solving techniques, no correlation between athletic self-esteem and shyness was found. These findings would suggest that shyness interacts with self-esteem, and that shy people, depending on the specific situation in which they are shy, and the way they deal with it, can exhibit different levels of self-esteem.

As with introversion, I expected that some of the shy participants in the present studies would have low self-esteem, but it was the fact that they were shy, and therefore might be interested in accessing a shy self-help Web site, that was of interest here. For that reason, I did not assess self-esteem.

Social phobia is defined as " a marked and persistent fear of one or more situations in which the person is exposed to possible scrutiny by others and fears that he or she may do something or act in a way that will be humiliating or embarrassing" (DSM-IV, 1994, p.416). Between 2% and 8% of adults meet the diagnostic criteria for social phobia (Beidel & Turner, 1998). Henderson and Zimbardo (2001) state that, since 1977, the Stanford Shyness Clinic has defined chronic shyness as "a fear of negative evaluation that was (a) sufficient to inhibit participation in desired activities and (b) that significantly interfered with the pursuit of personal and professional goals". These definitions both revolve around debilitating fear of scrutiny resulting from expected negative evaluations by others. The fear of the negative evaluation of others is common to everyone. Turk et al. (2001) propose that it is reasonable to expect that evidence from research on social phobia may be applicable to shyness (although with less extreme results). Both

social phobics and shy people experience the same physical responses (heart palpitations, sweating, trembling, and blushing), and the same kinds of fearful thoughts of negative evaluation or of doing something humiliating or embarrassing (Turner, Beidel & Townsley, 1990). Socially phobic people are more hampered in their social and occupational life than shy people; as well, the behaviour starts earlier and is more chronic (Beidel & Turner, 1998). Research findings on social phobia show that people in both categories share an inordinate expectation of negative evaluations by others, with social phobia being much more debilitating than shyness.

Chavira, Stein, and Malcarne (2002) found about one-half of the top 10% of shy university students were socially phobic. However, 18% of people who fell between the 40th and 60th percentile on shyness also were diagnosable as socially phobic. This shows that socially phobic students are found in the middle ranks of the shyness spectrum. Heiser, Turner, and Beidel (2003) tested 200 introductory psychology students using the 13-item Revised Cheek and Buss Shyness Scale (Cheek, 1983) and an unpublished personality questionnaire (Hyler, as cited in Heiser et al. 2003). They found only 3% of the 104 least-shy students to be socially phobic; however, 18% of the remaining 96 shy students were socially phobic. Overall, that would generate an estimate of 10.2% of the total sample being socially phobic. This led me to expect that about 10% of my participants might also be socially phobic, with the preponderance in the *shy* subset. Many of the truly socially phobic might not even volunteer to participate, but as this research is focussed on shy people, I did not assess social phobia.

Communication Apprehension

Shy people are averse to talking (Zimbardo, 1977), but is that all shyness is? Rapp (1999) states that “shyness” is an alternative label for “communication apprehension”. McCroskey (1970) coined the term “communication apprehension” based on a concept that Phillips (1968) had called a “broadly-based anxiety related to oral communication”. Later, McCroskey (1977; p. 78) defined communication apprehension as “an individual’s level of fear or anxiety associated with either real or anticipated communication with another person or persons”. McCroskey and Richmond (1982) specified this view of communication apprehension more clearly, and compared it to shyness, as presented by several researchers including Zimbardo, Ebbesen and Maslach, (1977), Girodo (1978), and Cheek and Buss (1979, unpublished report cited in McCroskey & Richmond, 1982). McCroskey and Richmond (1982) found that shyness and communication apprehension occurred in the same kinds of situations and had similar effects. They proposed that both shyness and communication apprehension would predict withdrawal and reduced talking. The difference between shyness and communication apprehension was in the proposed sources of the trait. McCroskey and Richmond saw shy behaviour developing from any of several “presumed causal foundations”: a state, a trait, a temperament, a genetic predisposition, or simply a choice for self-interest. “Communication apprehension”, however, had only social anxiety as a root cause. McCroskey and Richmond theorized that shyness was a genus, the tendency to be timid, with communication apprehension a species of timidity, the tendency to talk less.

In an effort to operationalize communication apprehension, McCroskey (1970) prepared four Likert-type instruments, The Personal Reports of Communication Apprehension (PRCA). Each is designed to measure this anxiety in a specific group: college students, grades seven and ten students, and students involved in public speaking courses. Interpersonal communication, small group communication, and public speaking situations were the foci of the initial items, 30 from a scale assessing “stage fright” (Gilkinson, 1942) and 46 written by himself and graduate students. The PRCA (McCroskey 1970) covered the three aspects of shyness discussed earlier: the affective (being nervous, tense, fearful), cognitive (anticipating expressing an opinion, anticipating public speaking, confused and jumbled thoughts, self-consciousness, authority figures), and behavioural (posture, loss for words, trembling hands, avoid public speaking).

Leary (1983b) argues that researchers should clearly distinguish between behaviour and subjective experiences. Leary (1991) categorizes the PRCA (McCroskey 1970) as being focused on the subjective, while the RCBS (Cheek & Buss, 1981) is seen to measure both subjective and behavioural facets of anxiety. However, Leary (1982), states that shyness presents a “special conceptual problem” due to the variety of definitions that cross the behaviour/ subjective experience boundary. The issue is confusing because of these varying definitions. I took the view that a definition of shyness must encompass affective, cognitive, and behavioural components.

When McCroskey and Richmond (1982) compared shyness and communication using the PRCA-College and a shyness questionnaire (SHY) that

they had developed earlier (McCroskey, Andersen, Richmond, & Wheelless, 1981), they found no relationship with shyness. In fact, their factor analysis showed that communication apprehension and shyness were “empirically distinct”. They did note, however, that this may not have been found if the Cheek and Buss (1981) scale had been used, as many of the items on the PRCA and the Cheek and Buss scales are very similar. It seems they may not have followed the caution articulated by Leary (1983b) to distinguish carefully amongst the aspects of the various constructs surrounding social anxiety.

The construct of communication apprehension, measured by the PRCA, appears to be overlapping with shyness, encompassing the affective, cognitive, and behavioural components of shyness, but the SHY shyness scale focussed only on diminished talking as the indicator. As a result, it is difficult to compare and contrast findings about communication apprehension and shyness.

Of the three components of shyness (affective, cognitive, and behavioural), (Crozier, 2005), the shyness scale, “SHY”, (McCroskey et al., 1981) focussed only on the behavioural, and a limited range of behaviours: self-perceptions of shyness, talkativeness, and quietness. The SHY scale does not measure most of the aspects of shyness. The PRCA (McCroskey 1970), the measure of communication apprehension, on the other hand, covered a wider selection of the three aspects of shyness. The PRCA would likely show high correlation with the later scale of Cheek (1983).

In conclusion, shyness and communication apprehension are related, but only within the behavioural aspect of shyness. Likely, most shy people, using the

RCBS scale (Cheek & Buss, 1981) would register as communication apprehensive with the PRCA scale. Communication apprehension overlaps with shyness, but is not related to this thesis, since communication apprehension focuses strictly on the individual's fearful response to public speaking, not the wider social arena feared by the shy person.

A Type of Attachment Bond.

This section examines the well-documented child developmental theory of attachment to assess if it can explain some of the causation and unfolding of shyness in adulthood. This is expected to inform our understanding of shyness more completely.

Parents and other caregivers typically, respond to a newborn's displays of distress to try to satisfy that infant's needs. Bowlby (1946) described a group of "affectionless" teen-aged thieves in the London Child Guidance Clinic. Statistically, these children made up 14 of the total of 44 thieves, but none of the 44 in a control group (non-thieves in the same institution). Bowlby (1946) hypothesized that lengthy mother-child separation in the early years was a factor in forming a child's character. Later, Bowlby formed a comprehensive "Attachment Theory", to explain these children's interpersonal behaviour, as well as describe the more general unfolding of the mother-child relationship in infancy (Bretherton, 1992).

Primarily relying on ethological work and methods, Bowlby (1958) distanced himself from the idea that all attachment develops from a single need (Bowlby, 1958; van der Horst, LeRoy, & van der Veer 2008; Bretherton, 1992). He

chose to use direct observation of infants and children, rather than adult memories of childhood as in psychoanalytical practice. Because so little evidence existed about human infants, he relied a great deal on the observations of animal researchers extrapolating to human infants. These observations could be best explained, he contended, by invoking at least five innate, genetically-based behaviours, implemented in the first year of a child's life. The positive responses, the sensitivity, of the "attachment figure" (the mother or father or caregiver) to these behaviours bring a feeling of security to the infant. These behaviours are sucking, crying, smiling, clinging, and following (Bowlby, 1958). Ainsworth (1967) in studying mother-infant interaction in Uganda, found the lack of such maternal sensitivity resulted in an insecure infant. Bowlby (1982) reported that infants, around 7 to 8 months of age, show a special attachment to caregivers by crying when they leave and being sad when they are absent. Many early researchers viewed this *attachment bond* as a unique relationship between child and primary caregiver in which the child seeks security and comfort from the relationship (Ainsworth, Blehar, Waters & Wall, 1978). Schaffer (1971) suggested that the crucial criterion for a person to be an object of such an attachment bond is that the infant does not consider this "attachment figure" to be interchangeable with any other figure. However, Main (1999) states that infants can and do form several contemporaneous attachment bonds with different persons but that there exists a hierarchy of preference in these relationships.

Ainsworth and Bell (1970) explored Bowlby's theory using the "Strange Situation" paradigm. This experimental scenario involves observing children

between the ages of 12 to 18 months responding to a series of eight 3-minute episodes, that start with a stranger, a mother and her infant, being together in a friendly way for three minutes. The stranger is then left alone with the infant; after this, the mother returns and interacts with the infant for three minutes; then the infant is left alone. Finally, the stranger returns for three minutes, then the mother comes in, and the stranger exits. Observations are made and recorded for each of the eight “acts”. From observing nine different behaviours of the children in these situations, attachment theorists have identified four styles of attachment, based on an ‘anxiety’ dimension (secure/insecure) and a ‘close bodily-contact’ dimension (Ainsworth & Bell, 1970; Ainsworth, 1979; Main & Solomon, 1986). These attachment styles, shown in Figure 3, have been studied and the “strange situation” results have been replicated worldwide (Main, 1996).

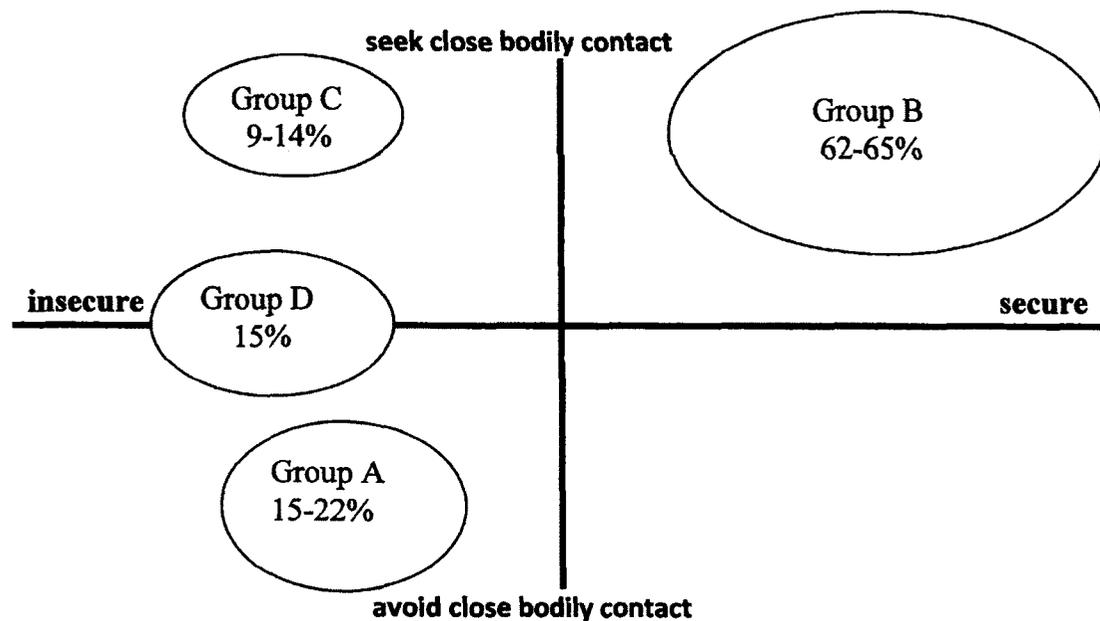


Figure 3. Attachment Styles (percentages from Prior & Glaser, 2006)

Figure 3 charts the classification of “Strange Situation” behaviour based on three studies involving a total of 4200 infants. The infants who were secure and seeking close bodily contact (Group B) would show distress and engage in less exploration on separation. They would seek proximity to mother on reunion and return to play (Ainsworth & Bell, 1970). Group B infants were the least angry and the most cooperative of the children (Ainsworth, 1979). The infants who were insecure and avoidant of bodily contact (Group A) would rarely cry upon separation, and would avoid (or ambivalently seek and then avoid) the mother in reunion (Ainsworth & Bell, 1970). They exhibited the most anger (Ainsworth, 1979). The mothers of Group A infants avoided bodily contact with the child and were more rejecting, more angry, showing less affect. The infants seeking bodily contact-insecure (Group C) showed pre-separation anxiety, were intensely distressed by separation and, although they sought the mother upon reunion, avoided contact or interaction with her. They could not be settled by the mother (Ainsworth & Bell, 1970). Those in Group D were disorganized/disoriented in their behaviour towards separation and reunion. They exhibited no obvious strategy (Main & Solomon, 1990), but would divert their attention to a toy upon reunion with their mother; cling to the mother, then push her away, then smile at the closed door; creep after the stranger, crying at the departure of stranger; often they “exhibited a hand to mouth gesture on reunion, resembling indecision, shame, or apprehension.” (Main and Solomon, 1986).

Main (1996) offers a description from “strange situation” procedures of maternal behaviour that feeds into the various attachment bonds in Figure 3. Infants

whose mothers sought bodily contact with them develop a secure view of the world, if the mother is tender and sensitive. If the mother is clumsy or nervous in holding the infant, unpredictable, and non-responsive in face-to-face interaction the infant becomes insecure (Group C). The mothers of Group A infants do not seek bodily contact resulting in the infant becoming insecure and avoidant of bodily contact. Main and Hesse (1990) suggest that parents who provoke both fear and reassurance cause the child to feel both secure and fearful, resulting in disorganization and confusion

Bowlby (1988) and Ainsworth (1989) theorized that attachment bonding is a part of adult life, as in a continuing relationship with a parent, or a new attachment with a spouse or a therapist. This has been tested empirically (Main, 1996) with an Adult Attachment Interview 20-item questionnaire (AAI, 2012). Trained researchers assigned the adults to various categories, based on their recall of their relationship with their parents. These categories were *secure*, based on positive childhood memories, *preoccupied*, based on closeness with parents but poor parental support, and *dismissing*, based in cold and unsupportive parents (Bartholomew, 1990). Bartholomew (1990) built on this AAI research, and theorized about adult attachment styles, categorizing them as in Figure 4.

		Model of Self	
		Negative	Positive
Model of Other	Positive	Preoccupied Overly dependent 14%	Secure (Group B) Comfortable with intimacy & autonomy 47%
	Negative	Fearful Socially avoidant 21%	Dismissing Denial of attachment 18%

Figure 4. Attachment Styles based on Bartholomew, 1990 (with percentages based on Bartholomew and Horowitz, 1991).

She argues, as did Bowlby (1982), that, based on the early experience with the mother, a person develops an inner working model of “the self” and “the other”. A *secure* adult attachment style (upper right quadrant) is seen to be based in a positive view of self and a positive view of the other (resulting from positive experiences with mother). This adult would have been classified in Group B by Ainsworth (1979).

Insensitive parenting paired with parental good intentions results in a positive sense of other but a negative sense of self, the *preoccupied* adult style, Group C in infancy. Insensitive parenting paired with a rejecting attitude on the part of the parent leads to a negative sense of other and a negative sense of self, the *fearful* adult style, Group A in infancy.

Bartholomew theorizes that the *dismissive* adult style results from negative infancy experience with a rejecting mother, but, for some reason, these infants form a positive self-image, as long as they do not interact with others. These would be the Group D infants in Main and Solomon (1986).

Bartholomew and Horowitz (1991) used a semi-structured interview to place 77 students in these four groups. They showed that self-report ratings, friend-report ratings, and 15 other measures supported the model. This framework can be connected to the group of shy people identifiable by the RCBS (Cheek, 1983). Shy people would perhaps be best classified in the *fearful* category as proposed by Bartholomew (1990). Bortnick, Henderson, and Zimbardo (2002) refer to the shy person's conflict between the desire to express a behavior and the fear of the consequence of expressing that behavior. This conflict is based upon the attachment history and attachment style of the person (Horowitz, Rosenberg, & Bartholomew, 1993). When the shy individual approaches a social situation, feelings of fear and negative predictions about the outcome of the social situation are prevalent; fear increases. Leaving the social scene relieves the pressure, but brings on self-loathing. This increases the person's expectation of rejection by others. Blaming others reduces the self-loathing, but also increases the expectation of rejection (Bortnick, Henderson, & Zimbardo, 2002). Bartholomew and Horowitz (1991), using the Inventory of Personal Problems Scale (Horowitz, Rosenberg, Ureno, Kalehzan, & O'Haloran, 1989) found that of the two avoidant styles, dismissive and fearful, only the *fearful* style (16 of 77 students) was consistently correlated with social insecurity (introversion) and lack of assertiveness. Similarly, Duggan and Brennan (1994) tested 676 introductory psychology students for sociability, shyness, and attachment style. They found a significant correlation ($r = .30, p = .001$) between the fearful group scores and their scores on shyness; the dismissive group correlation was non-significant ($r = .01$). Both of these groups had significant

negative correlations with the sociability scores. In that study, 24% reported themselves to be of the fearful attachment type. Bartholomew and Horowitz (1991) had reported 21%. These can be compared with the 42% of 2482 students who self-reported that they were presently shy (Zimbardo, 1989). It is likely that the *fearful* attachment style is a subset of the overall self-reported “shy” group. This seems to support the idea that trait shyness begins in childhood.

Henderson and Zimbardo (2001) delineate a model for shyness that parallels that for the *fearful* attachment style (Bartholomew, 1990), but the source of the expected rebuff or lack of a sensitive audience may come from another source than a distant, insensitive caregiver. So the fearful attachment style and Buss’s fearfully shy could be comparable, both arising from the physiological basis described by Kagan and Snidman (1999). The “shy” rubric however describes about twice as many adults as the “*fearful* attachment style” rubric. There appears to be a core of about 20% who are shy from early childhood but another 20% whose shyness depends on socio-environmental events that weaken a basically sound *secure* attachment style. Attachment style is likely one of several routes to shyness. I did not evaluate my participants for attachment style as the one group of interest, the “shy” subset, is a mixture of several categories of long-term and short-term socially avoidant people.

Summary of Shyness-Related Concepts

This section attempted to define shyness more carefully by discussing several related concepts. The concepts of introversion, low self-esteem, and social phobia are related to shyness. *Introversion* is moderately related to shyness.

However introverts can interact with other people without the fear, self-consciousness, and negative thoughts that shy people experience. The introverted person prefers solitary activities, while the shy person fears social activities.

Although the majority of shy people are also introverted, there are shy extroverts who are privately shy and publicly outgoing.

Shyness and *self-esteem* are negatively related overall, but the fearfully shy exhibit lower self-esteem than the self-consciously shy. Shyness involves the whole self, while self-esteem can vary depending on the situational context.

Socially phobic people are more hampered in their social and occupational life than shy people; as well, the behaviour starts earlier and tends to be chronic. Socially phobic students are found amongst the most shy but they exist also in the middle ranks of the shyness spectrum.

The above discussion suggests strongly that “shyness” is not an alternative label for “communication apprehension”. Communication apprehension is a set of behaviours in a particular context, social speaking, while shyness comprises cognitive, affective and behavioural aspects.

Lack of maternal sensitivity in child-rearing does seem to result in an insecure infant and a variety of troubling attachment styles in later life. Shy people would perhaps be best classified in the *fearful* attachment style. Attachment style is likely one of several routes to shyness.

The one group of interest to this thesis, shy people, is related to these various other definable groups, introverts, people with low self-esteem, the socially phobic, those apprehensive with respect to communication and those with a *fearful* adult

attachment style. Next, we examine the way that shy people view social situations with the aim of defining the goals of this thesis more clearly.

Shy People View Social Situations Differently

One of the aims of this thesis was to test if shy people view certain types of modification of Web pages differently than the general population. Research shows that shy people respond to a perceived audience differently from the general population (Turk et al., 2001; Henderson & Zimbardo, 2001). Socially anxious people anticipate being thought of unfavourably, even if the other person only glances at them (Leary, et al., 1988). This increased anxiety is supported by an augmented attention to negative signals. Using a signal detection theoretical analysis, Winton, Clark, and Edelmann (1995) showed that people with a high Fear of Negative Evaluation (FNE) score (Watson & Friend, 1969) had a negative response bias. Their threshold for rating a neutral face as negative (angry, sad, disgusted, contemptuous, or fearful) was lower than that of those who had low FNE scores. High FNE participants were more accurate at identifying negative faces, but they identified more neutral faces as negative as well, being less accurate in that situation. As well, Veljaca and Rapee (1998) found that people with higher social anxiety had a tendency to see negative valence even where it didn't exist, and also identified actual negative evaluations better than people with low social anxiety. These studies are not directly about shy people but shyness is closely related to fear of negative evaluation (Miller, 1995) and social anxiety (Henderson & Zimbardo, 2001).

Battaglia et al. (2004) found shy children who viewed a series of pictures of faces exhibiting a range of emotions (joy, fear, anger, disgust, sadness, surprise, and neutral) had difficulty deciphering the emotions expressed in the pictures. Later Battaglia et al. (2005) showed shy children, when exposed to expressions of joy, neutrality, and anger, demonstrated significantly less activity in the cortex when looking at the angry and neutral images. Battaglia and his colleagues concluded that shy children might go on “high alert” when presented with any face they can’t decipher and make use of more phylogenetically earlier fear responses based in the amygdala. It is also possible that “high alert” is a baseline for shy people. This explanation is supported by Pozzulo et al. (2005) who found that shy people who reported low arousal recalled more details about a “culprit” in a video of a staged theft. By contrast, in the same low arousal situation, the shy people recalled fewer of the general details of the crime scene. Those shy and least-shy students who reported high arousal displayed no differences in either type of recall. These differences in recall seem to support the view that shy people are more primed to see people’s faces than the least-shy. The shy people actually could have been at an optimum level of arousal to notice faces; their report of low arousal was due to the fact that “high arousal” is normal for them. The attention the shy people gave to the face of the culprit took away their attention from the other crime scene details. Faces seem to be more prominent to shy people than least-shy people. This enhanced prominence for faces, together with the negative affect shy people experience for a social situation, may make shy people evaluate a face on a Web

page more negatively than non-shy people. That contention was investigated in this thesis.

Simonian, Beidel, Turner, Berkes, and Long (2001) showed that children with extreme shyness, social phobia, had significantly poorer facial affect recognition skills than the normal population. Phan, Fitzgerald, Nathan, and Tancer (2006) also studying social phobic people, found that they demonstrated increased activity in the amygdala compared to controls when presented with threatening faces or social situations. The amygdala is a part of the brain involved in emotional responses as well as physiological and autonomic nervous system responses, breathing and heart rate. The above research on social phobia supports the view that shy people see and interpret statements and events differently from the general population. For shy people, people's faces are more salient, more prominent, and, representing the anticipation of a social encounter, more threatening than for the general population.

A Summary of Shyness

To summarize, "shyness" is generally, seen to have three components, the cognitive, the affective, and the behavioural. Shy people want to engage in social interaction, but are unable to, due to the cognitive and affective components. Shy people are fearful, some from a temperamental bias, some from a cognitive one. Physiological predispositions, as well as inter- and intra-cultural rearing practices could explain some of the sources of shyness. Shy people have more negative biases concerning the evaluations of social situations. For shy people, people's faces on a home page may be more salient, and more threatening, than for the general population. Once the faces are noticed, and evaluated negatively, the

shyness cycle, wherein the shy person peremptorily cuts off the feared situation, could cause shy people to evaluate the home page differently from the general population.

In the next section, I examine the framework within which I placed the research questions in this thesis, namely: i) do colour and photos affect the perceived trustworthiness and visual appeal of a home page? and ii) is there a difference between these evaluations by shy and non-shy people?

Trust, Visual Appeal, Colour, and Photos

Colour

There is evidence from artistic experience (Birren, 1982) and cross-cultural studies (Wiegersma & Van der Elst, 1988) that blue is a most preferred colour in general, with some variation individually and by region. Madden, Hewett, and Roth (2000) asked 253 undergraduate students from eight countries around the world to rate each of ten colours on a like/dislike scale. The average “liking” ratings ranged from 6.59 of 7 (blue in Austria) to 2.90 of 7 (orange in Colombia). Blue was rated highest in five of the eight countries, second highest in the other three. White was preferred most in two countries, second in three, and third in two. Interestingly, the lowest rating for white was fourth, in the United States. This study went on to search for meaning associations by rating the colours along twenty dimensions, using a seven-point semantic differential scale structure. A cluster analysis found that, consistently, blue, white, and green were rated similarly along the vectors “pleasant”, “calming”, and “gentle”.

Website designers pay attention to colour as shown by studies investigating colour variety between countries. Barber and Badre (1998) noticed that the colors of the national flags were usually used by governmental sites in all the countries they studied, with the exception of Brazil, which used a variety of bright colors. French sites made heavy use of blue, white, and red (the colors of the French flag) regardless of the domain, while Israeli and Lebanese Web sites frequently included the color green. Cyr, and Trevor-Smith (2004), in analyzing 30 municipal websites from each of Germany, Japan, and the United States, found that Germans preferred bluish purple (20% of the page) & blue (10% of the page), and that grey appeared most frequently on Japanese (10.4% of the page) and American (5.4% of the page) sites. Approximately three-quarters of the 90 sites used white as the background colour, grey coming next at 8.9%. Duncker, Theng, and Mohd-Nasir (2000) compared differences in color choices among Web sites created by students with various ethnic backgrounds. English students tended to use pastel colors, and frequently utilized grey, while Scandinavians favored dark hues. Students with a Jamaican background chose bright colors and used high contrasts. Black as a background color and colorful elements were frequently chosen by African students, while European and U.S. students tended toward bright backgrounds.

Complex results for colour preference and its relationship to other Web user characteristics, such as time spent on site, information retained, and trust are shown in several recent studies. Bonnardel, Piolat, and Le Bigot (2011) designed three versions of a 19-page Web site, one with a blue background, one orange, one grey. Each participant navigated one of these sites and wrote a letter to a friend about the

site. Both the orange and blue sites were rated significantly more appealing than the grey site, but only the orange site saw participants spend more time. Also, they recalled more information from the orange site than the other two. The aesthetic preference for blue did not seem to enhance these two variables as did the preference for orange. Interestingly, another study in the same investigation had shown blue to be the most liked relative to either an orange or grey home page. Cyr (2008) asked 223 Chinese, 230 Canadians, and 118 Germans to respond to the local version of the SonyStyle Web site and use it to choose a cell phone with the goal of buying it. The Canadian site used a large group picture, with red blue and white colours. The German site, predominantly white, displayed information simply and well-spaced, with no images of people. The Chinese site, predominantly blue, presented information in multiple locations on the page and used no pictures. Participants rated their perceptions of navigation design, visual design, information design, trust, satisfaction, and loyalty. A partial least squares structural analysis showed a significant relationship ($p < .001$) between visual design and trust only for the Chinese group. This connection was not found for the Canadian or German group. This lack of effect of colour supported Badre (2000) who, investigating localized Web sites, found that for Italian participants, the Italian localizations were preferable for navigation markers, but not for colour. Cyr, Head, and Larios (2010) modified the colour of the left navigation bar and the top graphic of localized versions of the SonyStyle website for Canadian ($n=90$), German ($n=90$) and Japanese ($n=90$) participants. Using a 13-item survey (containing five items for colour appeal and three for trust), they found, for each group, colour appeal was

significantly related to trust for the three colour variations, grey, blue and yellow. Noiwan and Norcio (2006) asked Thai and American students to rank blue, green, orange, red, violet, and yellow from most favourite to least favourite. Blue was most preferred by the two cultures but 30% of the Thai sample rated yellow 1st or 2nd, only 5% of the Americans did so. Violet was the least favourite colour for both groups. Using 7-point scale reports of usability for 12 colour combinations, both the Thai group and the American group rated the white text on yellow lowest for web appeal enhancement, web search enjoyment enhancement, web search interest enhancement, and readability categories. It is of interest that Sony as of January 02 2012 uses the same design and graphics for all of its European sites, only the languages are different. This design is primarily white with three rows of graphics/text boxes, primarily grey or black, with flashes of green, blue, and red. The present Canadian site is similar but requires more scrolling. The Chinese site requires no scrolling, and is mostly grey/black text on white, with some flashes of green, blue, and red. Sony appears to have focussed on the navigation design aspect and avoided the colour aspect perhaps due to the variable results shown between cultures.

Kim and Moon (1998) provided further evidence for the effect of colour on *trust*, by designing a specific interface to generate certain feelings in cyber banking customers. Using a 40-item, emotion-rating scale, 60 participants evaluated 26 interfaces. The participants were asked to draw a few randomly selected interfaces from memory, sort the 26 interfaces based on similarity of visual aspects, and explain their criteria for the groupings. A content analysis, based on these

qualitative findings, identified 14 salient design factors. One of them was colour. Continuing the study, 123 native Koreans evaluated twelve cyber-banking interfaces, chosen from the 26 in the second study. Factor analysis established a 7-factor “emotion space”, accounting for 80.6% of the variance. Attractiveness accounted for 50.0% of the variance, symmetry 11.7%, sophistication 12.8%, *trustworthiness* 8.2%, awkwardness 3.5%, elegance 2.4%, and simplicity 2.4%. It would appear that some combination of these variables could cause changes in certain emotions. Notably, for the study at hand, one of these emotions is trustworthiness.

For the final segment of the research, Kim and Moon (1998) used a 32-item subset of the original scale. Two interfaces were built, one with cool, pastel tones, pastel background, and low brightness, the other with bright, primary colours (blue, yellow, and red) and a bright background. The first resulted in increased reported feelings of trustworthiness towards the interface. In the introduction, Kim and Moon (1998) postulate that the interface has similar trustworthiness-altering effects to those shown to hold for the appearance of a potential human trustee.

Kim and Moon (1998) present several *caveats* to their work, in its real-life applicability to cyber-banking. However, one of these perceived drawbacks makes their findings very applicable to the study at hand. Participants in their research were asked to indicate their immediate feelings and, although this might not support extended cyber-banking trust, it does relate perfectly to the first stage of trust acquisition proposed by Sillence et al. (2004, 2006, 2007). Participants in our

studies were also asked to indicate their immediate feelings with respect to trust and visual appeal.

Gorn, Chattopadhyay, Yi, and Dahl (1997), attempting to more strictly control the bases for user decisions with respect to colour, based less on content, and more on the colours themselves, used a fake ad with a simple image of a solid color paint “swoosh” for their study. In order to avoid the content of the Web page overwhelming the first reactions of the Web site viewer, the third study in this thesis employed a Web page that was simple in design and offered little content.

Kim and Moon's (1998) results provide some evidence that colour may have an effect on trust. Their findings could be culturally specific, but they gave direction for some of the hypotheses in this thesis. Fogg, Soohoo et al. (2002), found that comments concerning “design look” aspect of “credibility” included references to “soothing green” versus “jarring purple”. However, Sillence et al. (2004) mention “lots of white, lots of pale blue”, and “insipid green”, as examples of design issues that could lead to mistrust of a Web page. It is not clear that a pastel shade of blue or green on a white background may be perceived as more trustworthy than many other colour combinations. Study 3 in this thesis examined that possibility.

Should we expect a difference between shy and less-shy people on the issue of colour preference? Ireland, Warren, and Herringer (1992) presented evidence that anxious people prefer pastel shades more than the non-anxious people. They tested 66 college undergraduates (30 high-anxiety, 36 low-anxiety), using six presentations of five shades per colour. Essentially, they asked each participant

“Which of these shades of [insert colour] do you prefer?” They found that the high-anxiety group preferred less saturated shades across the six colours (red, blue, yellow, green, brown, and purple) than did the low-anxiety group. They suggest this could be due to the less-arousing characteristics of pastels, or the fact that pastels are less flamboyant and allow the highly-anxious person to be more inconspicuous. Higher arousal, galvanic skin response readings (GSR), have been found for red stimuli over green, red over blue and yellow, and green over blue (Wilson, 1966; Jacobs & Hustmyer, 1974). Lichtlé (2007) found that individuals who seek a higher stimulation level found a saturated red hue to be more pleasurable and to generate a better attitude towards an advertisement than either a less-saturated red or blue. Noiwan and Norcio (2006) found a significant difference between the Thai and American groups in usability-reports for attention-drawing, web appeal enhancement, web search enjoyment enhancement, and web search interest enhancement for each of the 12 banner colour combinations. The Thai group rated all the reports significantly higher than the American group except in the case of attention-drawing, white text on yellow banners. Noiwan and Norcio (2006, p. 118) go on to speculate that the Thai group was more anxious because: the Thai were novice users, their culture might make them more anxious, and the researcher perceived that the Thai students exhibited more anxiety than the Americans. Elliott and Maier (2007) propose a model in which colour conveys information, within a context, specified by our learning and perhaps by our biology, particularly identifying a stimulus as hostile or hospitable (Elliott & Covington, 2001). Colour causes us to react, altering our emotions and behaviours (Valdez & Mehrabian,

1994) having to do with approach and avoidance (Elliott & Maier, 2007). This view of Web use, modified by anxiety levels, would be very applicable to the responses of shy and non-shy people to various Web page colour manipulations. Shy people may interpret colour differently from the general population, based on their physiological and/or learned predispositions.

So, in sum, although people prefer blue, green, and white (Wiegersma & Van der Elst, 1988; Madden et al., 2000), those seeking a higher arousal level prefer a saturated red and high-anxiety people prefer pastel colours over more saturated shades (Ireland et al., 1992). There is a great deal of variability with respect to the reaction to colour both within and between cultures, based upon learning and biology. As discussed earlier, anxiety and shyness are related (Kagan & Snidman, 1999; Heiser et al., 2002). Shy people are anxious, and so they might be expected to prefer pastels more than the least-shy population. Therefore, the effect of pastel and saturated colours on Web page trustworthiness and visual appeal was investigated here.

Photos

Some research indicates that the use of photos will increase trust in a Web site. Fogg et al. (2001) found that a Web page displaying an article was rated significantly more trustworthy with the addition of a formal photo of the author. However, the effect of a photo is not straightforward. Fogg et al. Also showed that *a casual photo* of the author of an article on the Web site *reduced* the mean trustworthiness score significantly. Steinbruck et al.(2002) found that, for participants using a fake on-line banking Web site for 15 minutes, a photo increased

trustworthiness over a site with no photo. A photo with a label identifying the person as a bank employee made the site even more trustworthy. This use of a photo is thought to replace the lack of face-to-face cues available through gaze and eye-contact (Steinbruck et al., 2002).

Riegelsberger, et al. (2003) also found complicated results for the effect of a photo on the credibility of an e-commerce Web site. They used four Web sites from each of three on-line vendors' categories (digital cameras, computer hardware, and flower services). Two of the sites in each category had been shown to have a good reputation, two a bad reputation according to reputation ratings from two online portals. They found that Web sites that are perceived as credible without a photo, gained no credibility when a photo was added. Less credible sites gained credibility from the addition of a photo. They proposed this interaction might occur because the photo, on certain sites, was seen as a manipulative tool. There was no main effect across all Web sites, so if the designer did not know if the site would be credible without a photo, the designer would be taking a chance on using a photo. Riegelsberger and Sasse (2002) used a thinking aloud paradigm, focussed interviews and content analysis to identify groups of on-line shoppers and their behaviours. They found four types of users: those with pre-existing high trust who were positively affected by the correct, "appropriate" photo; a more function-oriented group who mostly rejected photos as extraneous to the task; one group that did not normally shop on-line who gave the most positive reaction to the photos; another group were very risk-aware and saw the photos as manipulative. This last group tended to not trust on-line vendors, and as a result did not shop on-line. As

well, they did not believe the photographs were actually employees of the company. This subdivision of the sample is offered to show that putting photographs on a Web site might actually deter some of the target group. This group, with a tendency to not trust, could very well include some or all of the shy people in the studies at hand.

Shy people possess a sensitivity to notice faces (Winton et al., 1995; Pozzulo et al., 2005), and a tendency to assign negative valence to the faces (Leary et al., 1988, Simonian, et al., 2001); these characteristics could guide their interpretation of screen elements along the trust dimension differently from less-shy people. We could speculate that shy people, noticing items on a Web page differently from the general population, might reject a particular Web site that would be found quite trustworthy by the general population. We know that shy people fear social situations and the very anticipation of a social situation, so it may be expected that a photo inserted in a Web page may elicit this fear and cause a lower trustworthiness rating than that given by a non-shy group.

Photos can make a difference, increasing or decreasing trust. These differences may be specific to particular types of Web sites. The research shows a need to test photo-trust interaction in different types of on-line situations. There is also a need to segment the broad user market to identify how a photo may alter the perceptions of a specific customer. This thesis focussed on shy people, asking whether the trust engendered by a photo would differ for shy and non-shy people.

Trust and Visual Appeal

Trust and attractiveness (visual appeal) have been shown to be connected in bargaining games and evaluations of celebrities. Wilson and Eckel (2006), employing a money-based bargaining game (Berg, Dickhaut, McCabe, 1995) and 206 Introductory Economics students, found that attractive subjects were trusted at higher rates, i.e., “gained more money” (7.4%), in the first stage of the game. In the second stage, on the other hand, if the attractive partner did not meet expectations, they were penalized. Ohanian (1990) tested 108 undergraduate students on a Likert-type scale to validate a measure to rate the perceived trustworthiness, expertise and attractiveness of celebrities, and found a significant correlation between the “trustworthiness” component and the “attractiveness” component of the scale, indicating that visual appeal (attractiveness) and trustworthiness may share 18% of their variance.

The prominence of design "look and feel" (Sillence et al., 2007); or “Design Look” (Fogg, Soohoo et al., 2002) in the initial evaluation of the credibility of Web sites might be confounded by the visual appeal of the pages. Therefore, I examined how trust and visual appeal are related in the initial viewing of a Web page. This interest was supported by the findings of Lindgaard et al. (2006) who correlated visual appeal with seven design characteristics. There were high correlations ($r=.93$ to $r=.96$) between visual appeal and five of these characteristics: interesting/boring, imaginative/unimaginative, good/bad design, good/bad colour, and good/bad layout. Lindgaard et al. (2006) concluded that there was remarkable agreement on the visual appeal of the home pages. Lindgaard, et al. (2011) asked undergraduate

students to rate 50 home pages on visual appeal, trustworthiness, and usability. The correlation between visual appeal and trust in the initial evaluation was .62 ($p=.001$). The prominent aspects of the pages, in the first 500 milliseconds of exposure, were design, not information. As mentioned before, Sillence et al. (2004, 2007) presented a three-stage model of Web surfing trust acquisition, with first stage trust primarily resulting from design, and the design features focussing on visual factors (“busy layout”, “boring ...use of colour”, “small print”). Similarly, Fogg et al. (2002) found the judgements of nearly half of all health Web site consumers tested were influenced by the design of a site, including color schemes. Trustworthiness of a Web site is thus connected to the visual design aspect of a Web page.

The Visual Appeal Scale

Hassenzahl (2004) points to a large body of non-empirical literature about beauty in Human-Computer Interaction, noting that the particular words to be used to describe visual appeal is only in its beginnings. This is shown by the discussion in Lindgaard et al (2006) when they try to assimilate some of the words they used in their visual appeal study to those of Hassenzahl (2004). Is the English pair “interesting-boring” equivalent to German “lahm-fesselnd”? This cultural issue is discussed in Kim, Lee and Choi (2003) wherein the authors identify design factors which elicit particular emotional dimensions (using descriptors such as “bright”, “deluxe”, and “classical”). Lavie and Tractinsky (2004) developed a measurement instrument to measure the aesthetics of Web sites. Using a literature review and the advice of HCI researchers, web designers, interior designers and an architect, they found 41 adjectives about aesthetics. A three-stage method using large samples of

participants resulted in a two-factor structure for visual aesthetics. The authors labelled these “classic” and “expressive” aesthetics. The 10 adjectives retained in the final scale were *aesthetic, pleasant, clear, clean, and symmetric* for classic aesthetics, and *creative, fascinating, use of special effects, original, and sophisticated* for expressive aesthetics.

Statistics used in this Thesis

This section contains an overview of the decision process used in this thesis to choose the appropriate statistical methods for the studies. The tasks in several of the studies were related to ranking of Web pages rather than trying to establish how far they were apart. Non-parametric, also called distribution-free statistics, accomplished this very well. Non-parametric statistics change a data-set such as “65, 53, 34, 33, 25” to one made up of the ranks “1, 2, 3, 4, 5” before calculations occur.

All data analyzed in these studies comprised sums of scores on multi-item 7-point scales. There is some discussion as to the applicability of parametric tests to this type of data (Jamieson, 2004, 2005; Pell, 2005; Norman 2010), since the scores may not be interval measures or be normally distributed. This thesis took the position that the individual item response from “1” to “7” on a semantic differential scale was a simple ordinal figure, but, after being summed, over several items, the total for the scale could be treated as an interval number. The normality of variables was tested using standard skewness and kurtosis values. Skewness is a measure of the symmetry of the data distribution about the mean. High skewness values indicate a large number of very high or very low scores. Kurtosis is a measure of

the shape of the distribution, flat, bell-shaped, or very peaked. Absolute values over 2 were judged to indicate a significant deviation from normality (Cramer, 1997). Where the data were sufficiently normal, I employed parametric statistics, where not, I used appropriate non-parametric alternatives. I present here a short description of the non-parametric statistics for the reader who may not be familiar with these.

The Friedman Two-Way Analysis of Variance by Ranks (Friedman, 1937) is a repeated-measures non-parametric comparable to the parametric repeated-measures ANOVA. The scores are ranked by subject, across each row, and calculations based on the sums of these ranks are compared to the χ^2 distribution to identify significant differences across the conditions. If this test is significant, the Wilcoxon Matched-Pairs Signed-Ranks Test (Wilcoxon, 1945, 1949) can be used, similarly to the parametric paired-sample t-test, to search for the locus of the differences. That test ranks the absolute difference between the scores for each subject, allowing for ties. If there are 20 subjects, this will yield at most 20 differences and 20 rank positions. Some are positive, some negative. If there were 11 positive differences for example, there will be nine negative differences. The Wilcoxon W statistic equals the sum of the ranks of the lesser number of differences, and it is tested for significance against a probability distribution.

In the present studies, participants ranked web pages, and it was of interest to analyze the similarities of their evaluations. This was shown using Kendall's Coefficient of Concordance (Kendall & Babington Smith, 1939). Based on the same ranking as for the Friedman test, but comparing ranks between all of the pairs

of participants, Kendall's W ranges from 0 (no agreement) to 1 (complete agreement).

In Study 5, some between-subjects analysis was needed. This was accomplished by the use of the Mann-Whitney U statistic (Mann and Whitney, 1947) which ranks the two sets of scores along a single scale, then calculates if there are significant inequalities in the number of higher ranks in one set than the other.

Several studies focused on the rankings of Web pages, so no attempt was made to use parametric statistics. In others the distributions of response scores were analyzed for normality, employing skewness and kurtosis values with respect to their standard errors, and also comparing histograms to normal curves, as well as examining the Q-Q plots.

Throughout the thesis the tests all used an experiment-wise $\alpha = .05$. For multiple comparisons, this significance level was divided using a method proposed by Ryan (1960) that portions the "alpha" levels for multiple comparisons (probability of a Type 1 error) using the formula $\frac{2\alpha}{k(r-1)}$, where k = the total number of treatment levels, and r = the number of levels between the treatments being tested. This method allows for a variable significance level as the distance between means or ranks changes, generating a greater chance of finding a significant result. For example, for 10 samples, the significance levels required a range between .01 and .001.

Effect size was measured parametrically, using Cohen's d (Cohen, 1992), a statistic showing the number of standard deviations between the two means of the

samples. Cohen (1992) recommended that, when comparing independent means, 0.20 standard deviations would suggest a small difference, 0.50 a medium difference and 0.80 a large difference.

All data analyses were executed using SPSS 12.0. All correlations were made with the parametric Pearson Product-Moment statistic when the data distribution required and supported that measure. Otherwise, the non-parametric Spearman rank correlation was employed. All correlations were two-tailed.

Study 1

The purpose of Study 1 was to establish a *prima facie* ranking on trustworthiness for a set of health-related Web pages using an ad hoc 16-item trust scale based on the findings of Sillence et al. (2004, 2006, and 2007) and Fogg, Soohoo et al. (2002). The relative positions of these Web pages in this ranking, using that trust scale, were then compared to rankings on another two trust scales in Study 2.

As well, in order to guard against participants making ratings based on socially desirable responses, a caution mentioned by Sillence et al. (2004), the Marlowe-Crowne Social Desirability scale (Appendix N) was used.

In order to allow for more efficient statistical analysis in Study 1, a pilot study was conducted with two goals: i) to reduce an original set of 20 Web pages to eight, and ii) to select the most appropriate items for the ad hoc scale. A full account of the pilot study, summarized next, can be found in Appendix A.

Synopsis of the Pilot Study

First, I chose 20 home pages from a Google search on “healthy living” (Appendix A). This genre of Web sites was chosen because i) it is a topic of interest to most if not all people, whether they are shy or not and participant motivation could be expected to be high and equal for all participants; ii) it was a genre explored by Sillence et al. (2004, 2006, 2007) and Fogg et al. (2002), and therefore their findings with respect to design look factors would be applicable.

However, I wished to reduce the number of Web pages to increase the statistical power of the data analysis, since the more comparisons are made, the lower the power will be (Ryan, 1960). As well, I chose 16 words from the research of Sillence et al. (2004, 2007) and Fogg, Soohoo et al. (2002) for an ad hoc trust scale (Appendix B). Although these 16 items came from sound qualitative research, it was impossible to determine whether the words I chose from those studies were of equal weight vis à vis the concept of "trust". The pilot study, in which I correlated the rank orders between each scale item, allowed me to focus the scale on a single concept, thought to be "trust".

The rankings of these eight pages were correlated using the Spearman function, for each of the 16 items in the scale (Appendix D). It was thought that, if certain items generated similar rankings, they could be expected to tap into the same concept, "trust". As a result of this analysis, 11 items were chosen for the trust scale to be used in Study 1. A complete explanation of the selection process is given in Appendix E.

After being shown one of the 20 home page images ten participants, rated trustworthiness on the 16 item 7-point scale. The non-parametric Friedman test (Friedman, 1937) showed that the home pages exhibited highly significant differences (Appendix C). The Wilcoxon signed rank test (Wilcoxon, 1945) showed that significant differences existed between the three highest ranked Web pages and the three lowest-ranked pages. Eight pages (Appendix A) were selected, based on trustworthiness ratings: three rated as most trustworthy (the Natural Foods, Healthy Living, and Manitoba government sites), three rated as least trustworthy (the ArticleWorld, Healthy-Living, and Diabetic Gourmet sites) and two rated in the middle (the WeightWatchers and Healthy.net sites).

Formal Study 1

The purpose of this study was to establish a *prima facie* ranking on trustworthiness for the set of eight healthy living Web pages selected in the pilot study.

Method

Participants

Thirty-one participants, 15 male and 16 female, took part in this study. All were in their late teens or early twenties, recruited through an electronic bulletin (Appendix F) posted on SONA, the Carleton University Psychology Experiment Sign-Up System. All were recompensed with 1% added to their Psychology course mark. Each participant was tested individually in a session that lasted approximately 45 minutes. Shyness was not assessed.

Apparatus

The laptop computer used was a Toshiba Satellite A200 AH9, 1.73 GHz Intel CPU, with 2 GB RAM. The monitor was set at 32 bits, 1280 x 800 pixel screen resolution, 60 Hz refresh rate, with screen settings: Gamma 1.0, Brightness 0, Contrast 50. JPG images, at 720 x 563 pixel resolution in 32-bit colour, were presented as screen shots for each Web page as they would appear in an Internet Explorer browser window. The laptop screen was leaning away from the participants at an angle of approximately 30° to the vertical. The screen was approximately 60 cm from the participant. An application created in MediaLab v 2008.1.33 (Jarvis, 2008) was used to present Web pages and collect responses to the scales.

Materials

The materials included informed consent forms, debriefing forms, and pens. The eight Web pages selected from the pilot study (Appendix J and Figure 5 below) were used as the treatment levels of the independent variable. The rating instrument used (Appendix K) was the 11-item, 7-point semantic differential scale, from the pilot study. The labels in Figure 5 are the names of the sites and the Page # refers to their ranked order in the pilot study.

As well, in order to gauge whether participants could be answering based on some desire to please the researcher, or give a socially acceptable answer, the 13-item Marlowe-Crowne Social Desirability Scale (Appendix N) was used scale was administered. A non-significant correlation with the MCSD would indicate validity

for the 11-item trust scale, in that it is shown not to be simply a measure of a desire to answer in a socially desirable way.

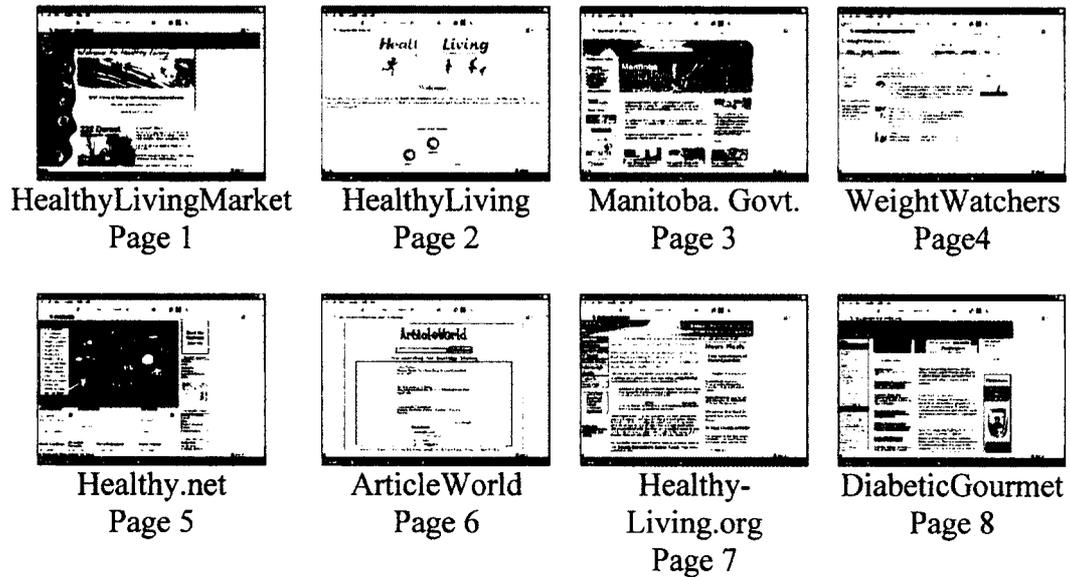


Figure 5. Eight Web Pages from Pilot Study, in rank order.

To guard against any confound due to demand characteristics each of the studies in this thesis used a precise script, with no intimation of the hypotheses available to the participants until the debriefing portion. “Demand characteristics” refers to the situation in which participants respond based on their perceptions of the experimenter’s wishes (Orne, 1962), resulting in attempts to confirm, or disconfirm, the researcher’s hypothesis (Nichols and Maner, 2008). The recruitment notices (Appendix F) and informed consent forms (Appendix G) were non-committal as to the purpose of the studies and the instructions were standardized by using computer-generated screens. The demeanour of the researcher was neutral. If participants asked for further information, I simply stated that I would answer those questions at the end of the session.

Design

This study involved a repeated measures design, with eight treatment conditions, the eight home pages from the pilot study. There were two within-subjects dependent variables, the trust scale resulting from data analysis in the pilot study and the social desirability scale.

Procedure

To begin with, the participants were welcomed and asked to sign a consent form (Appendix G). Then the 13-item Shorter Marlowe-Crowne Social Desirability (MCSD) scale was filled out, using paper and pen. Next, each participant was shown eight Web site home page images, in a randomized order to control for order effect. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants could take their time looking at the images, and that a smaller replica of the image would be presented below each of the statements to be rated. After seeing each page for as long as they wished, the participant clicked a "Continue" button, and, while still viewing a smaller image of the page, rated each page on 11 factors, one at a time, by clicking on a number between 1 and 7 on a semantic differential scale, displayed on the computer screen. The 11 items were presented in the same order to each participant. If there was a fatigue effect so that the 11th item was evaluated differently from the first, this effect would be equal over the pages. Since only the total score for each participant was analyzed, this was not expected to be an issue. The participants were told they could move the screen and/or the computer to a

more comfortable viewing position if desired. Finally, the participants were debriefed (Appendix I), thanked, and excused.

Results

The total score on the 11-item trust scale, for each of the eight Web pages, and each of the 31 participants, generated a 31 x 8 data matrix, each element an integer between 11 and 77 (11 rankings between 1 and 7). The scores were submitted to the non-parametric Friedman test (Appendix L). This generated a $\chi^2(7) = 95.733$, ($p < .001$), showing that the Web pages exhibited significant differences. The order of the rankings, is shown by mean ranks in Figure 6.

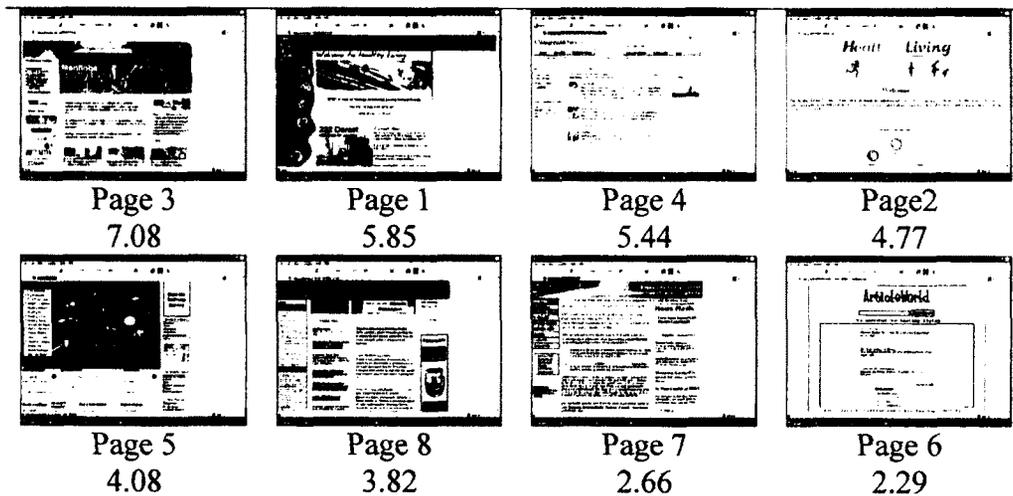


Figure 6. Web Pages from Study 1 with mean rank

Then, Wilcoxon matched-pairs signed-rank comparisons were made, to identify pages that were rated significantly more trustworthy than others (Appendix L). The pages were sorted according to their mean ranks from the Friedman test. Then, comparisons were made between the pages ranked farthest apart (Page 3 and Page 6), and continued until non-significant results were found. Then the next farthest apart were compared (Page 1 and Page 6) and so on. The Wilcoxon

comparisons showed several significant differences in trustworthy ratings as shown in Table 1.

Table 1

Friedman Ranks and Wilcoxon Results (row by column) for Trust Scale.

	Friedman	Wilcoxon Results					
	Mean Rank	Page 4	Page 2	Page 5	Page 8	Page 7	Page 6
Page 3	7.08	.004**	.000*	.000*	.000*	.000*	.000*
Page 1	5.85			.004*	.002*	.000*	.000*
Page 4	5.44			.006*	.002*	.000*	.000*
Page 2	4.77					.001*	.000*
Page 5	4.08					.001*	.000*
Page 8	3.82						
Page 7	2.66						
Page 6	2.29						

* $\alpha = .05$, controlling for experiment-wise error rate

An inspection of Table 1 shows that: i) Page 3 (the Manitoba government home page) was rated as significantly more trustworthy than six of the home pages; ii) Page 1, Healthy living Market, and Page 4, WeightWatchers, were significantly more trustworthy than four of the pages; iii) Page 6 (ArticleWorld) was significantly less trustworthy than all of the other pages, except for Pages 7 and 8.

Providing a different view of the Wilcoxon calculations, Table 2 shows the number of participants who preferred one Web page over another, when these were paired.

Table 2

Subjects preferring one page over another (Row Preferred over Column)

	Page 3	Page 1	Page 4	Page 2	Page 5	Page 8	Page 7	Page 6
Page 3		20	24	26	29	27	31	30
Page 1			18	20	21	23	28	28
Page 4				18	22	22	27	27
Page 2					18	20	24	27
Page 5						17	21	24
Page 8							20	21
Page 7								16

Thus, for example, 24 participants (77%) preferred Page 3 (Figure 7) to Page 4, 18 participants preferred Page 1 (Figure 7) to Page 4, and so on. This kind of information provides potentially more meaningful information to Web designers than conventional p-values as it provides more accurate details.



Figure 7. Three Most Trusted Web Pages.

Kendall's W was calculated to reveal the level of agreement in participants' rankings of the Web pages. It yielded a figure of .441, $p = .000$ (details in Appendix L), showing modest, but highly significant agreement among the participants. This indicated that the results were based on a consensus, rather than being weighted by a few of the participants.

Pearson product-moment correlations were calculated between the “corporate look and feel” item (Item 7) and the other 10 items for each of the eight Web pages, to see how this item fared vis-à-vis trustworthiness. Of the 80 correlations, 23 were significantly positively related. These are shown in Table 3. “Corporate look and feel” was positively correlated to several of the items that we assumed tapped into trustworthiness, especially, “professional”, “organized”, and “familiar”.

Table 3

Correlations of Item 7, “corporate look and feel” and other Items for each Web Page

Item	Page 3	Page 1	Page 4	Page 2	Page 5	Page 8	Page 7	Page 6
1: appropriate name								
2: clear layout						.406**		
3: navigation			.465**			.406**		
4: not boring				.536*				.734**
5: good colour		.395*					.529**	
6: print size								
8: professional	.416*			.476**	.574**	.415*	.822**	.458**
9: organized	.501**	.412*			.410*	.522**	.382*	
10: graphics	.369*						.516**	
11: familiar		.489**		.477**			.482**	

* p<.05, ** p<.01 (both 2 tailed)

The Table shows that item #7, “corporate look and feel”, correlated positively, with item #8 “professional” for trust scores for six of the Web pages (Pages 3, 2, 5, 8, 7 and 6. It was positively correlated with item #9 “organized” for five pages. The items related to “appropriate name” (item #1) and “print size” (item #6) showed no cases of significant correlation. I concluded that these participants viewed “corporate look and feel” as trustworthy.

In order to determine if the responses to the trust scale were biased by a desire to answer in a socially-acceptable manner, Pearson correlation calculations were

performed between the Marlowe-Crowne Social Desirability scores and each set of Web page scores. No significant correlations were found, indicating that the participants' responses were not based on any perceived socially more acceptable answers.

Discussion

Study 1 was successful in using the 11-item trust scale to discriminate along the trust dimension for the eight health-related Web pages. The rankings demonstrated a clear advantage for the Manitoba government and HealthyLiving Market Web page over the other pages, and a clear lack of trustworthiness for the HealthyLiving.org and ArticleWorld pages.

One interesting detail was seen in that Sillence et al. (2004, 2007) suggested that people viewed "corporate look and feel" as a *negative* indicator of trustworthiness, causing them to reject a Web site in an initial evaluation, but, in the pilot study (p. 74 and Appendix E), this item correlated *positively* with ten of the 11 items retained for Study 1. As in the pilot study, the *corporate look and feel* item was positively correlated in many instances. The highest rated pages were the Manitoba government site (Page 3), the Healthy Living Market site (Page 1), and the WeightWatchers site (Page 4), each rated as well-designed, exhibiting professional and corporate design look and highly trusted by these participants.

This sample of students did not find "corporate look and feel" to be a damaging consideration for initial trust in a Web page, in contrast to Sillence et al (2004, 2007) where this was associated with mistrust. That study used U.K. female participants, 41-60 years of age, and Fogg et al. (2002) used online participants

from across the U.S., of both genders, with an average age of 41.9 years. This study involved Canadian and foreign university students in their late-teens to early twenties, of both sexes. It is possible that age and culture affect the evaluation of corporate look and feel as it relates to trustworthiness of a Web page. As well, preferences change over time, and the earlier research was made five years ago.

Age and gender are important context variables, and these may explain why only some of the factors from the Sillence et al. (2004, 2007) list of design look words and phrases seemed to tap into the same construct as in Study 1. The middle-aged ladies in England quite likely noticed different aspects of the Web pages than the Carleton University undergraduates. Moreover, even when both groups rated the same factor, "corporate look and feel", they assigned a different valence to the concept, negative for one group, positive for the other.

Originally, I had been interested in web-based support for middle-aged women re-entering the work force as entrepreneurs, and I had intended to use participants who fell into this category. However, it became evident that accessing a large sample over several studies would be prohibitive, so I chose to use undergraduate students. These students were different from the groups studied by Sillence et al (2004) or Fogg et al. (2002) and therefore the results of this study, and the others in this thesis, may not be generalizable to other different populations. However, university students are possible users of a web-based shy support site, and so a reasonable, real-life, test group, to guide web design principles for shy undergraduates and undergraduates in general. Greater generalizability will come by replication with other samples and other types of web pages.

In a study by Nisbett and Wilson (1977a) shoppers were asked to rate the quality of four pairs of stockings, which were in fact identical. When asked to explain their preference, the shoppers invented reasons even though a pronounced bias was evident. When this was mentioned as a possible reason for the choice, most participants denied this could be a cause. Asking people why they chose a Web site may produce similar outcomes. People may not know why they have made their choices, but they nevertheless offer reasons. This may be why some of the items in Sillence et al. (2004) did not correlate highly with others.

The sample size and different populations used in this thesis extended the validity of the model proposed by Sillence et al. (2004, 2007) and made it more generalizable. The use of the Marlowe-Crowne Social Desirability scale also addressed a possible limitation mentioned by Sillence et al. (2004), that their participants, in a group setting, might have made socially desirable responses.

The purpose of this study was to establish a *prima facie* ranking based upon the trustworthiness research of Sillence et al. (2004, 2007) for the set of eight Web pages selected in the pilot study. This was accomplished. This was necessary in order to provide stimuli for Study 2, wherein two measures of trust would be compared to provide some criterion validity for the adapted Jian scale.

Study 2

This study was designed to test the criterion validity of the Jian Trust Scale (Jian et al, 2000) which had not been shown previously to be suitable for evaluating initial trust in a Web page. This was therefore assessed here by ranking trustworthiness for the eight Web pages from Study 1 using the Jian Trust scale. In

order to compare it with another scale that has also been widely used in a Web context, relevant sections of the McKnight Trust scale were also selected and used to measure trust. Both were slightly adapted to suit this study as described in the Materials section below.

The Jian scale (pp. 24ff), was selected as the measurement instrument for trustworthiness in this thesis because it focused on general trust in a non-human entity. By comparing results from these two scales, this study aimed to ensure that the adapted Jian scale would be an appropriate and valid instrument to use in the forthcoming studies concerning initial trust in a Web page.

The items in the two trust scales used a 7-point ruler anchored to the left by "1" and "Disagree", to the right "7" and "Agree". Care was taken to avoid bias but if there was some order effect, but this was expected to be towards the left, the "disagree" end (Friedman & Amoo, 1999). Such a bias would be uniform across the scale and factor out statistically, since all pages would experience the effect equally

Method

Participants

Forty-four participants, 28 female and 16 male, were recruited from the same age group and source as before. All were recompensed with 1% added to their Psychology course mark. No evaluation of shyness was necessary. Each participant was tested individually in a session that lasted approximately 45 minutes.

Apparatus and Materials

The apparatus was the same as in Study 1 (p. 80). The materials included informed consent forms, debriefing forms, and pens. The eight Web pages selected

in Study 1 were used as the experimental stimuli, as well as the adapted Jian scale and the adapted McKnight trust scale.

The Jian scale was adapted by replacing the word "system" by "Web site" in nine of the 12 scale items. Since I was testing first-time experience with the selected Web pages, I omitted the statement "I am familiar with the system" completely. The other three statements were modified slightly, resulting in an 11-item scale (Appendix J).

The statements in the adapted Jian scale, as in the original, were worded so that, for five items, "trust" anchored at the left end of the scale ("1"), and for six items, "trust" was anchored to the right end of the scale ("7") to control for any tendency to respond in a particular direction.

Eleven statements, relevant to initial trust in Web pages, from the *Trusting Beliefs* and *Trusting Intentions* sections of the original McKnight scale were modified. These items were anchored with "Agree" to the right-hand end as in the original scale and shown in Figure 8 below. Any bias in response pattern was expected to be towards the left, the "disagree" end and was expected to factor out statistically, since all pages would experience the effect equally.

Each item was presented separately on the computer screen as below (Figure 8).

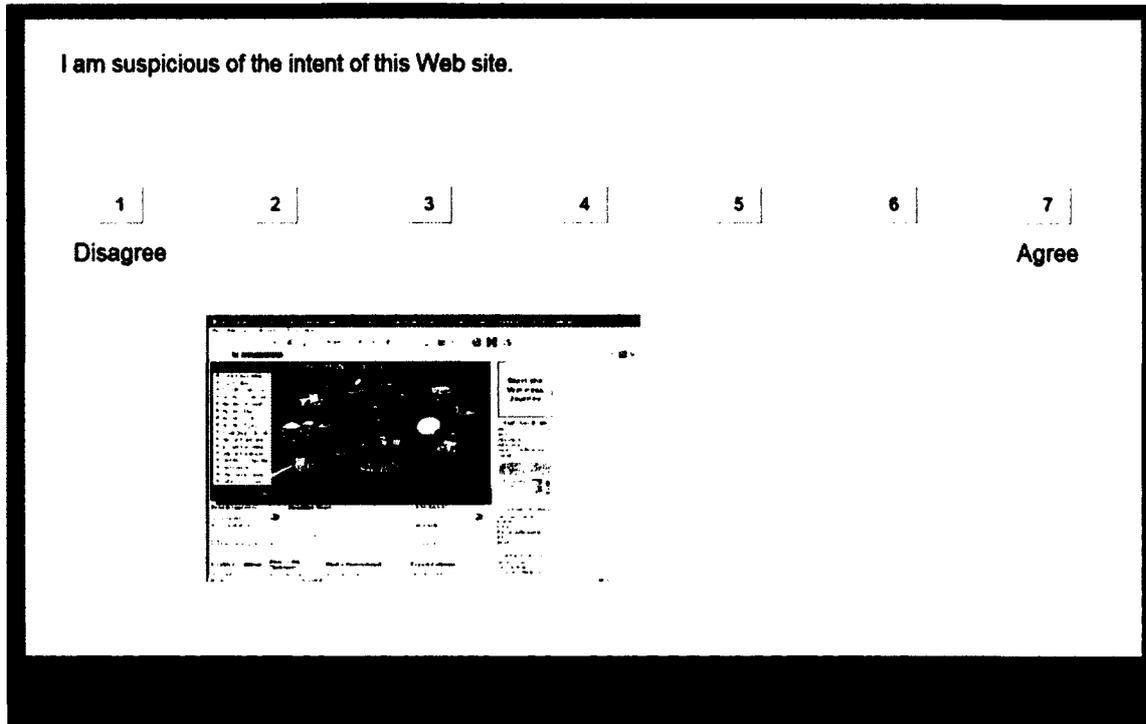


Figure 8. Example Item in 22-item Trust Scale.

A complete description of the adaptations can be found in Appendix M.

These two adapted scales were presented as a single 22-item combined trust scale (Table 4 below).

Table 4.

Combined 22-item Scale.

Item	Jian items
1	This Web site is deceptive.
2	There is something underhanded about this Web site
3	I am suspicious of this Web site's intent.
4	I am wary of This Web site.
5	If I used this Web site, its actions would have a harmful or injurious outcome.
6	I am confident in This Web site.
7	This Web site provides security.
8	This Web site has integrity.
9	This Web site is dependable.
10	This Web site is reliable.
11	I can trust This Web site.
McKnight items	
12	I believe that this Web site would act in my best interest.
13	If I required help, this Web site would do its best to help me.
14	This Web site is interested in my well-being, not just its own.
15	This Web site is truthful in its dealings with me.
16	I would characterize this Web site as honest.
17	This Web site would keep its commitments.
18	This Web site is sincere and genuine.
19	I would feel comfortable using this Web site.
20	I can always rely on this Web site.
21	I feel that I could count on this Web site.
22	I would not hesitate to use this Web site.

As well, in order to gauge whether participants could be answering based on some desire to please the researcher, or give a socially acceptable answer, the 13-item Marlowe-Crowne Social Desirability Scale (Appendix N) was administered. A non-significant correlation with the MCSD would indicate validity for the 11-item trust scale, in that it is shown not to be simply a measure of a desire to answer in a socially desirable way.

Design

This study employed a repeated measures design with a single experimental procedure and eight Web pages as the treatment levels. There were three within-

subjects variables, the two trust scales and the social desirability scale. The order of the presentations of the stimuli varied randomly for each participant. The order of the items in the scale was the same for each participant. Since the data analyzed were the total score by participant, any fatigue effect wherein the 22nd item was evaluated differently from the first, would be equal over the pages, and therefore, was not expected to be an issue. Responses were made on a computer monitor as in Study 1.

Procedure

To begin with, the participants were welcomed and asked to sign a consent form (Appendix G). Then the 13-item Shorter Marlowe-Crowne Social Desirability (MCSD) scale was filled out, using paper and pen. Next, the participants were shown each of the eight Web pages on a computer monitor and rated each using the 22-item combined trust scale. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants could take their time looking at the images, and that a smaller replica of the image would be presented below each of the statements to be rated. Finally, the participants were debriefed, thanked and excused.

Results

The results are divided into five sections. First, (1) data exploration, followed by (2) the correlation of the Jian scores and the McKnight scores, for each of the eight Web pages. Next, (3) the Jian results and (4) the McKnight results were assessed separately using the Friedman and Wilcoxon matched-pairs statistics. Then, (4) the rank orders of the eight pages, based on the Jian scale, the McKnight

scale, and Study 1 were compared. Next, (5) the Jian scores and scores from the Marlowe-Crowne Social Desirability Scale were correlated to test whether the participants' responses were based on perceived socially more acceptable answers. Finally, (6) a correlational test was made of the unidimensionality of trust and distrust.

Data Exploration

First, in order to ascertain the appropriate statistics to employ, the distributions of responses for the trust scores as well as the MCSD scores for the eight Web pages were analyzed, attending especially to skewness and kurtosis values with respect to their standard errors.

Of the 16 trust variables, six displayed a normalized skewness value over 2, indicating a significant deviation from normality. Two of these variables also yielded extreme kurtosis values. As a result, the data were analyzed non-parametrically. Non-parametric analyses are appropriate for a study such as this, since the goal is to establish a rank order on the basis of design look/initial trust, and not to be more specific with respect to distance between the pages.

Correlation of the Jian and McKnight scores overall

Next, the Jian total scores and the McKnight total scores for the 44 participants were non-parametrically correlated, using the Spearman rank-correlation coefficient, for each of the eight Web pages. Table 5 shows the results.

Table 5

Spearman Correlations between Jian & McKnight Scores by Web Page

	Page 4	Page 6	Page 2	Page 1	Page 3	Page 8	Page 5	Page 7
r(43)	.934***	.886***	.894***	.940***	.883***	.934***	.942***	.936***

*** p<.001 (2-tailed)

As the Table shows, all these correlations were statistically highly significant, indicating that the Jian and McKnight scales generated very similar ratings for each of the eight Web pages with respect to trust.

The two sets of scale results were also evaluated separately, in order to compare the ranks assigned by each to the eight Web pages

The Jian Scale

Based on the 11-item Jian scale, the scores were analyzed using the non-parametric Friedman test. This generated a $\chi^2(7) = 117.018$, ($p = .000$), showing that the Web pages exhibited significant differences. The Wilcoxon matched-pairs signed rank test was then used to determine the loci of significant differences amongst the eight Web pages. These are shown in Table 6. The significance level for the set of comparisons was set at $\alpha = .05$. Using the Ryan technique, this required a p-value of 0.001786 for the Page 3 versus Page 6 comparison and 0.003125 for the Page 4 versus Page 2 comparison.

Table 6

Friedman Ranks and Wilcoxon Results for Jian Scale.

	Ranks	Wilcoxon Results						
		Page 4	Page 8	Page 1	Page 5	Page 2	Page 7	Page 6
Page 3	6.91	.000*	.000*	.000*	.000*	.000*	.000*	.000*
Page 4	5.84					.002*	.001*	.000*
Page 8	4.77							.000*
Page 1	4.64							.000*
Page 5	4.33							.000*
Page 2	3.99							.000*
Page 7	3.70							.000*
Page 6	1.82							.000*

* $\alpha = .05$, controlling for experiment-wise error rate

As the Table shows, Page 3 was rated significantly more trustworthy than all of the others, and Page 6 was rated significantly less trustworthy than all the other pages. The other pages fell between these two extremes, with Page 4 higher-rated than three other pages.

Providing a different view of the Wilcoxon calculations, Table 7 shows the number of participants who preferred one Web page over another, when these were paired.

Table 7

Wilcoxon Results. Row Preferred over Column

	Page 3	Page 4	Page 8	Page 1	Page 5	Page 2	Page 7	Page 6
Page 3		29	37	38	36	36	37	44
Page 4						35	33	40
Page 8								42
Page 1								39
Page 5								34
Page 2								35
Page 7								33

Thus, for example, 29 (66%) of participants preferred Page 3 (Figure 9) to Page 4, 37 participants preferred Page 3 (Figure 9) to Page 8, and so on.

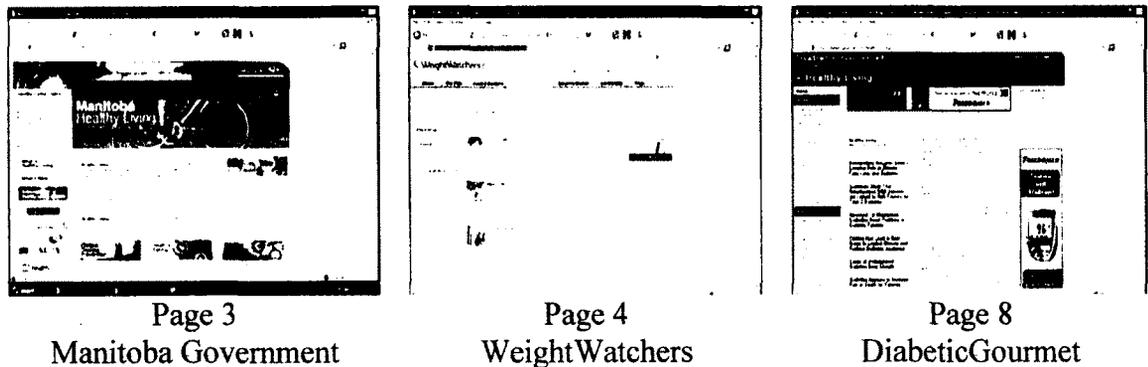


Figure 9. Three Most Trusted Home Pages.

Figure 9 illustrates the three most trustworthy Web pages based on the Jian trust scale.

Kendall's W was also calculated, and yielded a figure of 0.380 ($p=0.000$) showing a significant amount of agreement among the participants. This was a useful statistic as it indicated that the results were not being weighted by a few of the participants, but that the ratings were based on a consensus.

The McKnight Scale

Using the 11-item McKnight scale, the scores were submitted to a non-parametric Friedman test. This generated a $\chi^2(7) = 120.412$ ($p < .001$), showing that the Web pages exhibited significant differences.

The Wilcoxon matched-pairs signed rank test was then used to establish the particular significant differences amongst the eight Web pages. These are tabulated in Table 8 below.

Table 8

Friedman Ranks and Wilcoxon Results for McKnight Scale

	Ranks	Wilcoxon Results (p-values)						
		Page 4	Page 8	Page 1	Page 2	Page 5	Page 7	Page 6
Page 3	6.95	.000*	.000*	.000*	.000*	.000*	.000*	.000*
Page 4	5.90				.002*	.001*	.001*	.000*
Page 8	4.85							.000*
Page 1	4.36							.000*
Page 2	4.35							.000*
Page 5	3.99							.000*
Page 7	3.80							.000*
Page 6	1.80							.000*

* $\alpha = .05$, controlling for experiment-wises error rate

These results, similar to those for the Jian scale, indicated that Page 3 rated significantly more trustworthy than all of the others, and that Page 6 rated significantly less trustworthy than all the other pages. The other pages fell between these two extremes, with Page 4 better-rated than four other pages. Page 5 and Page 2 switched positions compared to the order generated by the Jian scale, but this difference in ranks was not statistically significant.

Kendall's W yielded a figure of 0.391 ($p < .001$) showing a significant amount of agreement among the participants, indicating that the results were not being weighted by a few of the participants, but that the ratings were based on consensus.

These rank orders were to be used, along with the rankings by the Jian scale, to show the degree to which the two scales rankings correlated, which would indicate a measure of criterion validity.

Correlations of Jian and McKnight trust scales

The rank orders from Study 1, Study 2 Jian scale, and Study 2 McKnight scale are shown below (Table 9).

Table 9

Friedman Rank Orders Studies 1 & 2

Study 2 Jian scale	Friedman rank	Study 2 McKnight scale	Friedman rank	Study 1	Friedman rank
Page 3	6.91	Page 3	6.95	Page 3	7.08
Page 4	5.84	Page 4	5.90	Page 1	5.85
Page 8	4.77	Page 8	4.85	Page 4	5.44
Page 1	4.64	Page 1	4.36	Page 2	4.77
Page 5	4.33	Page 2	4.35	Page 5	4.08
Page 2	3.99	Page 5	3.99	Page 8	3.82
Page 7	3.70	Page 7	3.80	Page 7	2.66
Page 6	1.82	Page 6	1.80	Page 6	2.29

Table 9 shows the Jian and McKnight scales agree on the placement of the first four pages and on the relative position of six of the eight pages. Pages 2 and 5 changed order, but this difference was not statistically significant. There is agreement across all three scales on the placement of Pages 3, 7, and 6.

The similarities of the ranks from Study1 and Study 2 are clear, but to assign some statistical measure to this common-sense observation, the rank orders were correlated using Spearman (non-parametric) correlations. The results were as tabulated below (Table 10).

Table 10

Spearman Correlations between rank orders Studies 1 & 2

	Study 1	Study 2 Jian	Study 2 McKnight
Study 1	1.000	.786 *	.810*
Study 2 Jian		1.000	.976***
Study 2 McKnight			1.000

* p< .05 level *** p<.001 level (both values 2-tailed)

These correlations indicate a strong similarity with the rankings from the ad hoc scale in Study 1, and confirm the notable similarity between the rankings assigned by the two trust scales in Study 2.

In order to determine if the responses to the Jian trust scale were biased by a desire to answer in a socially-acceptable manner, Spearman correlation calculations were performed between the Marlowe-Crowne Social Desirability scores for each of the Web pages (Table 11). No significant correlation was found. This would indicate that the participants' responses were not based on any perceived socially more acceptable answers.

Table 11

Spearman rank-correlation coefficient of MCSD and Jian trust scores

Marlowe-Crowne Social Desirability Scale							
Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8
-.131	-.125	-.217	-.040	-.058	.135	.094	-.021

Correlations of 10 Items in Jian Scale with Item #11

To test the contention by Jian et al.(1998) that trust and distrust are opposites, not different factors, Pearson Product-Moment Correlations were made of each of the items and the “I can trust this Web Site” item, Item #11, for each of the eight Web pages. Thirty-nine of the 40 correlations were significantly positive for the

five trust-oriented items (confident, security, integrity, dependable, and reliable, and thirty-nine of the 40 correlations were significantly negative for the five distrust-oriented items (deceptive, underhanded, suspicious, wary, and harmful). (Appendix O).

Discussion

Both the scores and the ranks for the Jian scale correlated significantly with the McKnight scale within each Web page.

This study contributed to the assessing of criterion validity of the adapted Jian Trust Scale, specifically in reference to initial trust in a Web page. It also showed that design look was a very real a component of the formation of initial trust in a Web site by the very similar rankings given to the eight Web pages using the design look based scale from Study 1 and the two dedicated trust scales. The Jian scale was designed to measure initial trust in Web pages per se, and did not include items connected to design look. The correlations between the ranks from Study 1 and those from Study 2 support that there is relationship between these two constructs, but they still remain unique concepts. Trust and visual appeal were investigated further in Studies 3 and 5. Suffice it to say, Study 2 demonstrated that the adapted Jian scale was appropriate to serve as a valid instrument to evaluate initial trust in a Web page in the following studies.

There were some differences in the ranks assigned to various pages by Study 1 and the Jian and McKnight scales. These differences are likely due to the clear focus in the Jian and McKnight scales on trust-related words, such as “confident”, “reliable”, and “sincere”, whereas the scale in Study 1 used items having to do

with design look, e.g. "clear layout", or "good colour". This suggests that design look, which feeds into initial trust, is still different from the kind of initial trust, as measured by the Jian and McKnight trust scales. A test of this would be to replicate the two studies with a repeated measures design, so the responses could be correlated to identify the shared relationship between perceived design look and perceived trustworthiness. This could contribute to the discussion concerning the degree to which design look contributes to initial credibility judgements. Recall that Sillence et al. (2004, 2007) found 94% of the factors relating to initial rejection of a Web site had to do with design, whereas Fogg et al. (2002) attributed 46% to design look. However, because the purpose of this thesis was to study the relation between trust and shyness, it was beyond the scope here to conduct such a study.

Study 2 supported a relation between trust and visual appeal. Lindgaard et al. (2006) and Lindgaard et al. (2011) found high correlations between visual appeal and several design-look factors, such as design, colour, layout, and graphics, which have been shown in Study 1 and Study 2 to be highly correlated with the initial trustworthiness of a Web page. The Study 1 scale, made up of design look items, correlated highly with the Jian trust scale ($\rho=.786$) and with the McKnight trust scale. ($\rho=.810$) in Study 2. So design look correlated with visual appeal as well as with trustworthiness; therefore visual appeal can be said to be correlated with trustworthiness.

The results of Study 2 supported Jian et al. (1998) in their position that trust and distrust are opposite extreme of a single dimension, and not two separate facets as presented by Lewicki et al. (1998).

Study 3

As discussed earlier (p.62ff) there are large individual variations in colour preferences. This applies to Web design as well. Colour seems to convey information within a context, specified by our learning and perhaps by our biology.

For these reasons, the purpose of Study 3 was to examine the effect of colour variation on the trustworthiness and visual appeal of a Web page, on a general sample as well as on a group of shy people.

Gorn et al (1997) used a simple image of a solid color paint “swoosh” for their study attempting to focus the participants less on content, more on the colours themselves. For the same reason, this study employed a Web page that was simple in design. The archetype stimulus Web page for Study 3 (Page 2, “Healthy Living.org”) met this criterion and offered little content placing focus on the colours.

Kim and Moon (1998) found that cool colour tones with a moderate pastel main colour resulted in increased reported trustworthiness towards the interface. Fogg, Soohoo et al. (2002), found that comments concerning the “design look” aspect of credibility included references to “soothing green” versus “jarring purple”. Sillence et al. (2004) mentioned “lots of white, lots of pale blue”, and “insipid green”, as examples of colours that could lead to mistrust in a health-related Web page. Taken together, these findings led to Hypotheses 1 and 2 predicting that for a general population, varying the colours used in a home page would alter ratings of trustworthiness (H1) and of visual appeal (H2). The colours used were saturated and pastel versions of blue, green, red, purple and yellow.

Based on Study 2 which showed that visual appeal can be said to be correlated with trustworthiness, Hypothesis 3 predicted that the trust scores would be highly correlated with the visual appeal scores.

To measure the perceptions of visual appeal for the various Web pages, this study used statements based on nine of the adjectives from Lavie and Tractinsky (2004). The phrase "*special effects*" was omitted since it did not appear to be appropriate as the Web pages used were static, with no "special effects". It was the goal to have a sufficient number of items on the scale to establish a reasonable expectation of reliability. Masters (1974) has shown that, given a 9-item scale with seven categories, we can expect a reliability coefficient of approximately .75.

Ireland et al (1992) showed that high-anxiety participants' preferred less saturated colours. Madden, Hewett, and Roth (2000) found that blue, white, and green were consistently rated similarly along the vectors "pleasant", "calming", and "gentle". Shy people are anxious, and so they might be expected to prefer pastels more than the saturated shades and more than the non-shy population. Therefore, it was proposed that shy people would rate the pastel shaded pages more trustworthy (H4) and more visually appealing (H5) than the same page with saturated color.

Hypotheses 6 and 7 predicted that shy people would rate the pastel shaded pages more trustworthy (H6) and more visually appealing (H7) than non-shy people. Finally, Hypothesis 8 and 9 predicted that shy people would rate the saturated-colour pages less trustworthy (H8) and less visually appealing (H9) than the non-shy people.

Table 12 presents these nine hypotheses.

Table 12

Five Hypotheses for Study 3

Hypothesis 1	For a general population, varying the colours used in a home page would alter ratings of trustworthiness.
Hypothesis 2	For a general population, varying the colours used in a home page would alter ratings of visual appeal.
Hypothesis 3	The trust scores would be highly correlated with the visual appeal scores.
Hypothesis 4	Shy people would rate the pastel shaded pages more trustworthy than the same page with saturated color.
Hypothesis 5	Shy people would rate the pastel shaded pages more visually-appealing than the same page with saturated color.
Hypothesis 6	Shy people would rate the pastel shaded pages more trustworthy than the non-shy people.
Hypothesis 7	Shy people would rate the pastel shaded pages more visually-appealing than the non-shy people.
Hypothesis 8	Shy people would rate the saturated-colour pages less trustworthy than the non-shy people.
Hypothesis 9	Shy people would rate the saturated-colour pages less visually-appealing than the non-shy people.

Method

Participants

84 Participants, 47 female and 37 male, were recruited from the same age group and source as before. The participants were recompensed with 1% added to their Psychology course mark. Each participant was tested individually in a session that lasted approximately 45 minutes.

Sample Size

Since hypotheses H3 through H6 dealt with trust and visual appeal evaluations by shy and non-shy participants, I required a sufficiently large sample to select 20 "shy" and 20 "non-shy" participants for these analyses. In 2007, 1038

Carleton University Introductory Psychology students who filled out the RCBS scale demonstrated a range of scores between 13 and 62, with 25% below 28 and 25 % lying above 39. I expected that, if I tested 80 participants, the highest and lowest 25% on the RCBS would be included in these two groups.

Apparatus and Materials

The apparatus was the same as in Study 1 (p.80). The materials included informed consent forms, debriefing forms, and pens. As well, ten Web pages were constructed, using KompoZer v. 0.7.10 (<http://www.kompozer.net/>), to be used as the treatment levels of the independent variable. HealthyLiving.org (Page 2), from Study 2 was used as the basis for the variations (Figure 10).

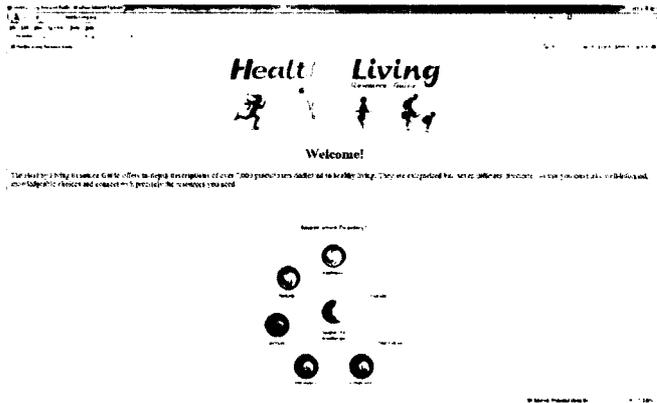


Figure 10. Original Page 2 “Healthy Living” Home Page

This Web page was selected as there was high agreement on its ranking in Studies 1 and 2 (4 of 8 in Study 1, 6 of 8 on the Jian scale in Study 2, and 5 of 8 on the McKnight scale in Study 2) and it was not most or least trusted . As well, I chose to use a simple page to avoid the content overwhelming the viewers’ first impressions. The page has a light-hearted graphic of children playing, with a set of navigation buttons in a circle at the bottom of the page.

The colours chosen were saturated and pastel examples of blue, green, red, purple and yellow, with the accurate RGB values given in Table 13 below. All colours (Appendix P) had Munsell scale equivalent measures except for pastel yellow. All graphics were manipulated to display only one colour apart from the text which was always black against a white background.

Table 13

RGB definitions for the 10 colours in Study 3.

	Blue	Green	Red	Purple	Yellow	Pastel Blue	Pastel Green	Pink	Pastel Purple	Pastel Yellow
Red	000	000	193	128	255	166	152	255	186	255
Green	083	107	000	062	220	189	199	126	127	255
Blue	138	060	032	117	051	215	147	147	162	160

An example of a stimulus page is found in Figure 11,



Figure 11. Amended Healthy Living Page (Blue)

The 20-item, 7-point semantic differential scale, (Appendix Q) comprised the adapted Jian Trust scale and the 9-item Lavie and Tractinsky Visual Appeal scale (pp74ff), comprising five items from the *classical* sub-scale, four from the *aesthetic* sub-scale. As well, the participants filled out the RCBS shyness scale (Appendix R).

Design

This study used a repeated measures, randomized block factorial design including one within-subjects independent variable, on-screen characters (colour, background), with 10 levels; and one between-subjects independent variable, shyness, with two levels (shy, non-shy). The participants were told to look at the Web page for a few seconds, then click a “Continue “ button, which would take them to a series of 20 statements, one at a time, to be rated on a 7-point semantic differential scale. They were told there would be a smaller version of the Web page, in the lower portion of each screen, accompanying the items so memory was not an issue. After responding to an item, the next item would appear, accompanied by the small Web page image. Both the presentation of the stimuli and that of the scale items were randomized in MediaLab 2008 to control for any order effect. The Agree/Disagree anchors were reversed for 11 of the 20 items to control for any tendency to respond in a particular direction.

Procedure

Participants were first asked to sign a consent form (Appendix G). Then, they viewed each of the 10 Web site home page images, which appeared as if they were viewed in an Internet Explorer browser, and rated each site on the 20-item scale. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants could take their time looking at the images, and that a smaller replica of the image would be presented below each of the statements to be rated. Finally, each participant filled

out the 13-item RCBS Shyness scale (Cheek, 1983), and was then debriefed (Appendix I) and excused.

Results

The results are divided into six sections. First, (1) data exploration, followed by (2) the comparisons between pages based on trust scores, and (3) the comparisons between pages based on visual appeal scores. Next, (4) correlations between trust and visual appeal scores overall are presented, followed by (5) the trust and visual appeal comparisons between pages for the shy and non-shy groups, and, finally, (6) the correlations between trust and visual appeal for the shy and non-shy groups.

Data Exploration

First, the distributions of responses ($n=84$) for the trust scores as well as the visual appeal scores for the 10 Web pages were analyzed, attending especially to skewness and kurtosis values with respect to their standard errors.

Of the 20 trust and visual appeal variables (trust and visual appeal for 10 Web pages) only the trust variable for the pastel yellow page (Page #6), displayed a normalized skewness value over 2, indicating a significant deviation from normality. The kurtosis values yielded no such concerns. These tests indicated that only one of the 20 data sets was significantly non-normal.

Investigating further, SPSS 12 generated Q-Q plots to chart actual scores against a normal distribution. If the data were normal, the plot would be distributed close to the straight 45° line. The plots for the 10 pages, both the trust and the visual appeal variables, all appeared to be close to the line. Especially of interest was the

plot for the pastel yellow page (Page #6), shown in Figure 12 below, because of its high skewness value. It also appeared to be a plot of a reasonably normal data set.

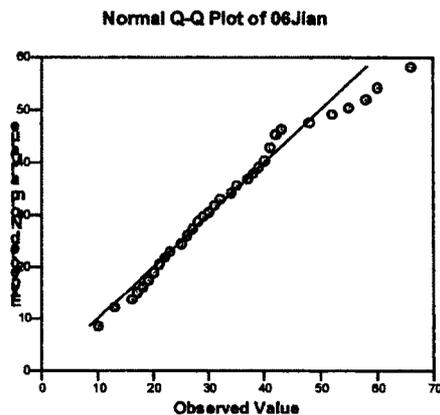


Figure 12. Q-Q plot for Page #6, Pastel Yellow. Jian scores.

All the variables for the 10 pages appeared to be similar to the data plot displayed in Figure 12, and therefore to be sufficiently close to normal to allow for parametric analysis. As well, the 840 scores across all 10 pages, for both the trust and visual appeal scales, appeared to be quite normally distributed. Therefore, I reported the parametric analysis fully, noting the non-parametric results only in the case of the trust variable for the pastel yellow page (Page #6).

The same normality tests, for the shy and non-shy samples, showed skewness issues with the pastel yellow (Page#6) for both trust and visual appeal data, and for the trust data for pastel purple (Page #5). Again, I have reported the parametric analysis fully, noting the non-parametric results only in the case of the trust variables for pastel purple and pastel yellow, and the visual appeal variable for pastel yellow.

Comparisons of Pages based on Trust scores

To test Hypothesis 1, predicting that, for a general population, varying the colours used in a home page would alter ratings of trustworthiness, a repeated measures one-way ANOVA with 10 levels of colour was computed for the trust variables. It was significant, $F(6.051, 502.268) = 22.953, p < .001$ using the Greenhouse-Geisser correction. Forty-five t-tests for related samples were then calculated to identify the loci of significant differences (Appendix S). Table 14 shows those significant differences, as well as the mean and standard deviations for the trust scores for the 10 pages.

Table 14

Jian Trust Scale t-tests

	M	SD	05 Pastel Purple	10 Yellow	06 Pastel Yellow
Blue	47.58	12.25	4.91*	5.17*	8.80*
Green	46.49	11.12			8.36*
Purple	46.31	11.49		3.44*	8.63*
Pink	46.06	13.39			8.52*
Pastel Blue	45.52	10.81			9.21*
Pastel Green	45.44	11.55			7.89*
Red	44.62	13.31			6.53*
Pastel Purple	43.44	11.59			7.45*
Yellow	43.13	12.84			6.36*
Pastel Yellow	34.88	11.71			

* $\alpha = .05$, controlling for experiment-wises error rate

The t-tests, showed that i) all pages were rated significantly more trustworthy than the pastel yellow page; ii) the blue page was rated more trustworthy than the yellow page and the pastel purple page; iii) the purple page was rated more trustworthy than the yellow page.

Figure 13 shows screen shots of the pastel yellow, yellow, and pastel purple pages.

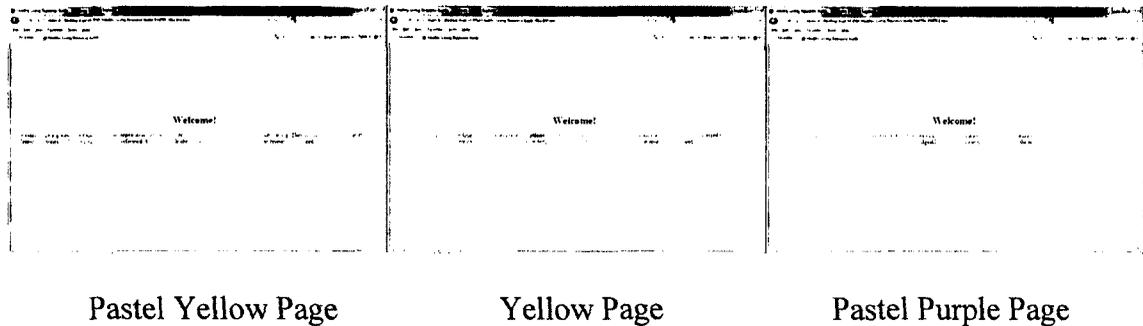


Figure 13. Three Home Pages.

A measure of effect size, the magnitude of the differences, can be obtained by taking the difference in means and dividing by the pooled standard deviation for the two sets. For the blue page and the pastel yellow page this yielded a difference of 1.06 standard deviations, considered to be a large effect (Cohen, 1992). For the yellow page versus the pastel yellow page the figure was 0.675, a medium to large effect.

A post hoc check of the raw data showed that 43 participants (51%) rated the blue page as more trustworthy than the red page and 33 (39%) rated the red page as more trustworthy, with eight ties.

Supporting these findings, the non-parametric Friedman test on the Jian Trust Scale scores generated a $\chi^2(9) = 120.744$ ($p < .001$). Subsequent Wilcoxon matched-pairs signed-rank comparisons (Appendix T) based on the Friedman rank order, for the pastel yellow and pastel purple pages, supported the parametric findings.

These findings thus supported Hypothesis 1, that varying the colours in a home page would alter the trustworthiness ratings. Since the basic Web page was so simple, lacking in much content, and many of the effect sizes were medium to large, these findings are noteworthy.

Comparisons of Pages based on Visual Appeal scores

To test Hypothesis 2, predicting that, for a general population, varying the colours used in a home page would alter ratings of visual appeal, a repeated measures one-way ANOVA with 10 levels of colour was performed on the visual appeal scores. The ANOVA was significant, $F(6.947, 576.635) = 37.447, p < .001$ using the Greenhouse-Geisser correction. Forty-five t-tests (Appendix U) were calculated, to determine the loci of significant differences. Table 15 shows those significant differences, as well as the means and standard deviations for the visual appeal scores for the 10 pages.

Table 15

Visual Appeal Scale t-tests

	M	SD	05 Pastel Purple	10 Yellow	06 Pastel Yellow
Blue	35.92	9.19	3.675*	5.906*	11.336*
Pink	35.63	10.12		4.128 *	11.264*
Green	35.48	8.64		3.973*	11.331*
Pastel Blue	35.50	8.39	3.503*	5.126*	12.232*
Purple	35.23	7.63		4.112 *	11.532*
Pastel Green	34.98	9.35		3.981*	11.077*
Red	34.32	9.15			9.340*
Pastel Purple	33.38	8.38			10.437*
Yellow	31.81	9.33			7.803*
Pastel Yellow	23.40	8.83			

* significant at $p < .05$ level, controlling for experiment-wise error rate

The t-tests showed that i) all pages were rated significantly more visually appealing than the pastel yellow page ii) the yellow page was rated significantly *less* visually appealing than six pages; iii) the blue page was rated more visually appealing than the pastel purple page, iv) the pastel blue page was rated more visually appealing than the pastel purple page.

A measure of effect size was calculated in the same way as for the trust scores. For the blue page and the pastel yellow page this yielded a difference of 1.39 standard deviations, a large effect (Cohen 1992). For the yellow and pastel yellow page the figure was 0.93 (a large effect), but for the blue versus the pastel purple page and the yellow pages, the effect size figures were 0.29 (a small effect) and 0.44 (a medium effect). These figures show the pastel yellow was by far the least visually appealing of all the pages, and that the effect of blue on the visual appeal of the archetype page was not large when compared to eight of the nine other pages.

Supporting these findings, the non-parametric Friedman test on the visual appeal scores generated a $\chi^2(9) = 170.119$ ($p < .001$). Subsequent Wilcoxon matched-pairs signed-rank comparisons (Appendix V) based on the Friedman rank order, for the pastel purple page, supported the parametric findings.

These findings supported Hypothesis 2, stating that varying the colours in a home page would alter the visual appeal evaluations. Since all of the effect sizes between the Pastel Yellow page and the other pages were large, these findings are noteworthy.

Correlations between Trust and Visual Appeal Scores Overall

Several methods were used to test Hypothesis 3 predicting that the trust scores would be highly correlated with the visual appeal scores. First, the 10 Web pages were ranked according to the trust and visual appeal scores by the means (Table 16). These ranks were then correlated. Second, all 840 trust and visual appeal scores were correlated, and finally, trust scores were correlated with the visual appeal scores for each of the pages. These three methods yielded different views of the relation between trust and visual appeal.

Table 16

Ranks of 10 Web pages by means for Trust and Visual Appeal

	Trust			Visual Appeal	
	Mean	SD		Mean	SD
Blue	47.58	12.25	Blue	35.92	9.19
Green	46.49	11.12	Pink	35.63	10.12
Purple	46.31	11.49	Green	35.48	8.64
Pink	46.06	13.39	Pastel	35.50	8.39
			Blue		
Pastel	45.52	10.81	Purple	35.23	7.63
Blue					
Pastel	45.44	11.55	Pastel	34.98	9.35
Green			Green		
Red	44.62	13.31	Red	34.32	9.15
Pastel	43.44	11.59	Pastel	33.38	8.38
Purple			Purple		
Yellow	43.13	12.84	Yellow	31.81	9.33
Pastel	34.88	11.71	Pastel	23.40	8.83
Yellow			Yellow		

The possible range for the means was 11 to 77, based on the 7-point 11-item trust scale. The blue page was ranked most appealing as well as most trustworthy,

the pastel yellow least appealing and trustworthy. Using these ranks, trust and visual appeal were strongly related ($r(8) = .988, p < .001$).

The second comparison between trust and visual appeal scores used all 840 as a group. This yielded a much lower Pearson parametric correlation ($r(838) = .601, p < .001$). This indicated that the relation between trust and visual appeal scores was such that they both resulted in very similar orders for the means or ranks, but that individually, there was more variability between the scores.

To further test how much the trust and visual appeal ratings overlapped, the trust scores were correlated with the visual appeal scores for each of the pages, with results as shown in Table 12 below.

Table 17

Correlations between trust and visual appeal scores by colour.
The order is by size of correlation.

	Pearson (parametric)
Blue	.645***
Pastel Green	.596***
Pink	.581***
Red	.571***
Yellow	.565***
Pastel Blue	.561***
Green	.554***
Purple	.545***
Pastel Yellow	.523***
Pastel Purple	.438***

*** $p < .001$, 2-tailed

As the Table shows, the Pearson correlations ranged from .438 (pastel purple) to .645 (blue), with a median of .563 (R^2 between 19% and 42%). Because the data for the pastel yellow page was very skewed, the Spearman correlation was applied. It supported the parametric test ($\rho(82) = .472, p < .001$). Therefore, it was safe to

conclude that the trust and visual appeal scores resulted in very similar mean orders, but that, within the individual pages, there was a great deal of unexplained variance. That was why the mean orders were quite highly correlated, but the 840 scores were less well correlated.

With respect to the 84 participants as a whole, H3, predicting that the trust scores would be highly correlated with the visual appeal scores, was partially supported. The correlations were significant, but ranged by colour, an issue which will be taken up in the discussion.

With respect to this set of stimuli, visual appeal and trust shared about 30% of the variance, with some colours, such as blue, being more closely associated, some others, such as pastel purple, less so. This wide range of correlations, gives a more accurate picture of the complex relationship between trust and visual appeal in the initial view of a Web page than that found when simply the mean orders were tested.

Trust and Visual Appeal Comparisons for Shy and Non-Shy

The next set of hypotheses concerned the ratings by shy and non-shy participants to the various Web pages. Therefore, the participants were divided into two groups of 20 using the RCBS scale (Cheek, 1983), “shy” (scores between 41 and 56, median = 44.5) and “non-shy” (between 16 and 25, median = 20.5). There were 11 males and 9 females in the shy group, 8 males and 12 females in the non-shy group. Comparing these cut-off scores to the 2007 Carleton Psychology undergraduate RCBS scores, 21% of those students had scores at or below 26, 25% had scores at or above 41. There were 7 of 1038 (0.7 %) of undergraduates with

RCBS scores higher than 56, and 14 (1.3%) with scores lower than 16. Both of these sets of scores were seen to be reasonably normal, after calculating skewness and kurtosis figures, and examining their histograms and Q-Q plots. These scores were significantly different using the independent sample t-test; $t(31.923) = 20.250$ ($p < .001$). The Mann-Whitney U statistic ($U = .000$, $Z = -5.427$, $p < .001$) confirmed this difference. It was reasonable to use the two groups of shy and non-shy people.

The 10 trust and visual appeal variables for each group were examined for normality. The non-shy trust variable for the pastel purple page, and both variables for the non-shy pastel yellow page violated assumptions of normality. Therefore, parametric assessment of these variables was accompanied by the equivalent non-parametric test.

When looking at the trust and visual appeal scores from the shy and non-shy groups (Appendix W), it is of interest that, in 19 of 20 cases, the maximum scores for each page were higher for the non-shy group and in 18 of 20 cases the minimum scores were lower for the non-shy group. In general, the non-shy group exhibited a wider range of scores than the shy group. They were more likely to use the whole scale than were the shy group. This supports the view of the cautious shy person, guarded even in assigning ratings on a trust scale.

The findings that non-shy participants rated the blue and pink pages more trustworthy than the shy group, and that the mean scores for each page were slightly higher for the non-shy group, initiated analysis of the 200 scores for these two groups (20 participants and 10 pages). These data sets were highly skewed, the shy to the right, showing a tendency to rate lower on the scale, the non-shy to the

left, showing a tendency to rate higher on the scale. Because of the large skewness values, the 200 scores from each group were subjected to the non-parametric Mann-Whitney two-sample rank sum test. Median trust scores in the shy and non-shy groups were 42 and 50; the distributions in the two groups differed significantly (Mann-Whitney $U = 13219.5$, $p < .001$ two-tailed). Median visual appeal scores in the shy and non-shy groups were 32 and 34.5; the distributions in the two groups differed significantly (Mann-Whitney $U = 16402.0$, $p < .001$ two-tailed). These results indicated that the shy group rated the pages more conservatively than the non-shy group, both for trustworthiness and visual appeal.

It was possible that the significant differences for the blue and pink pages were sufficiently large to affect the total sample of 200 scores, so an analysis was done of the eight other pages. These data sets were sufficiently normal to employ an independent-sample t-test. Even though, individually, none of these eight pages showed significant differences, taken as a group, the differences were highly significant. The trust means for the non-shy ($M=48.7$) were higher than the shy ($M=42.5$), $t(306.404)=4.277$, $p < .001$. The visual appeal means for the non-shy ($M=33.92$) were higher than for the shy (31.7), $t(297.272) = 1.990$, $p < .05$. These results indicated that the shy group rated the pages more conservatively than the non-shy group, both for trustworthiness and visual appeal. A measure of effect size was calculated for both trust and visual appeal. For trust, this yielded a difference of 0.48 standard deviations, a medium effect (Cohen 1992), for visual appeal 0.22, a small effect.

The first set of comparisons tested Hypothesis H4, that the shy group would rate the pastel shaded pages as more trustworthy than the saturated page of the same hue. Therefore, the shy group was tested on evaluation of the pastel version of a colour versus the saturated version, i.e. pastel blue versus blue, pastel green versus green, and so on, using the matched-pairs t-test, with the significance level set to $\alpha=.05/5$. The means are shown in Figure 14 below. The means are based on the 7-point scale for 11 items, generating scores ranging from 11 to 77.

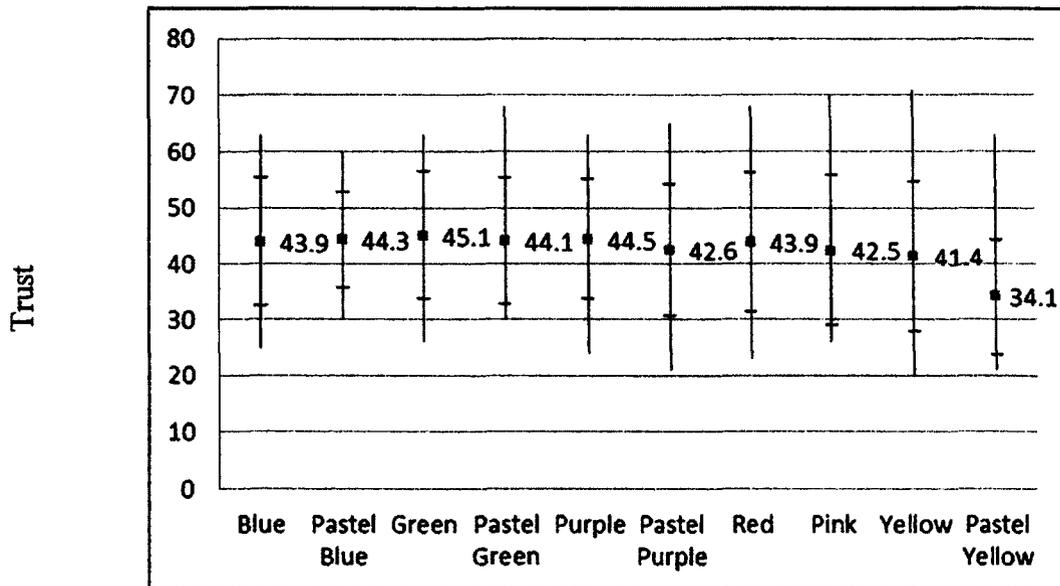


Figure 14. Shy Trust Means for 10 Web Pages. Each vertical line represents the range of scores. The horizontal bars mark one standard deviation.

As the figure shows, all but one of the trust score means were very similar, ranging from 41.4 to 45.1. Only the yellow/pastel yellow looked possibly significant. Analysis using matched- pair t-tests showed no significant differences of the pastel page over the corresponding saturated-colour page. Yellow was the only instance in which the pastel colour was given a lower score than the saturated for the shy group ($M=41.35$ vs. $M= 34.05$, $t(19) = 3.407$, $p < .01$, also supported by a

Wilcoxon test, $p < .001$). However, this was in the opposite direction to the prediction of Hypothesis 4. Hypothesis 4 was thus not supported.

Similar t-tests were computed to test Hypothesis 5, predicting that shy people would rate the pastel shaded pages more visually appealing than the same page with saturated color. The means are shown in Figure 15 below.

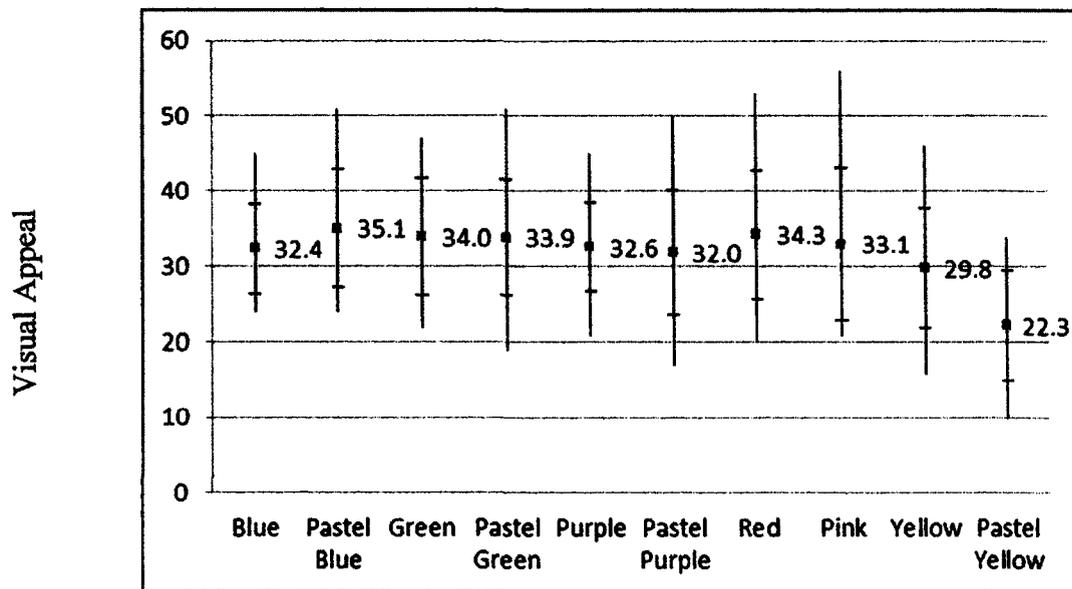


Figure 15. Shy Visual Appeal Means for 10 Web Pages. Each vertical line represents the range of scores. The horizontal bars mark one standard deviation.

As the figure shows, the pastel blue page appeared to have a higher average score than the blue page, the red higher than the pink, and the yellow higher than the pastel yellow page. However, analysis using matched- pair t-tests showed no significant differences of the pastel page over the corresponding saturated-colour page. However, the shy group found the yellow page more visually-appealing than the pastel yellow page ($M=29.8$ vs. $M= 22.3$, $t(19)=4.130$, $p<.001$, again in the opposite direction to that predicted by Hypothesis 5. Therefore, Hypothesis 5 was not supported.

The next set of shy/non-shy comparisons tested Hypotheses 6, that shy people would rate the pastel shaded pages more trustworthy than the non-shy people. To test this, evaluations of the pastel coloured pages were compared by an independent samples t-test, the significance level set to $\alpha=.05/5$. The results were not significant. Thus Hypothesis 6 was not supported. The only significant difference was that the Non-shy group found the pink page more trustworthy than the shy group ($M=55.10$ vs. $M=42.45$, $t(38) = 2.839$, $p<.01$), which was not predicted.

Similar t-tests for Hypothesis 7, that shy people would rate the pastel shaded pages more visually-appealing than the non-shy people found no significant t-tests. The hypothesis was thus not supported.

The next set of shy/non-shy comparisons tested Hypotheses 8, that shy people would rate the saturated-colour pages less trustworthy than the non-shy people. This was again tested by independent samples t-tests, the significance level set to $\alpha=.05/5$. These showed that the Shy group found the saturated blue page significantly less trustworthy than the non-shy group ($M=56.35$ vs. $M=43.90$, $t(38) = 3.212$, $p<.01$). Thus Hypothesis 8 was supported.

Similar t-tests for Hypothesis 9, that shy people would rate the saturated-colour pages less visually-appealing than the non-shy people found no significant t-tests, and the hypothesis was not supported.

Correlations of Trust and Visual Appeal for the Shy, Non-Shy, and Middle Groups

Hypothesis 3, stating that trust scores would be highly correlated with the visual appeal scores, was also tested separately for the shy and non-shy groups.

Trust/Visual Appeal correlations were made within the shy and non-shy groups, for each of the pages (Appendix X), and for all 200 scores generated by each group.

For the shy group’s scores, the Pearson product-moment correlation was $r(198) = .683, p < .001$; for the non-shy group’s scores it was $r(198) = .729, p < .001$. Recall that the correlation for the complete 84 participant sample was $r(838) = .601, p < .001$. Therefore, within each group, the trust and visual appeal scores shared more of the variance (47% and 53% on the parametric correlations) than was evidenced over the complete 84-participant sample (36%).

Table 18 shows the correlations between trust and visual appeal scores for the shy, non-shy, and the middle group for each of the 10 pages. The only non-significant findings in the table were in the middle group for which the shared variance ranged from 5% (pink) to 30% (blue).

Table 18

Pearson Correlations for Trust and Visual Appeal for each page within Shy, Non-shy, and middle groups.

	Shy n=20	Middle n=44	Non-shy n=20
Blue	.677***	.551***	.780***
Green	.566**	.533***	.600**
Pastel Blue	.612**	.515***	.646**
Pastel Green	.708***	.511***	.632**
Pastel Purple	.622*	ns	.602**
Pastel Yellow	.681***	.409**	.707***
Pink	.696***	.321*	.734***
Purple	.806***	ns	.699***
Red	.627***	.390**	.844***
Yellow	.682***	.485**	.630**

* $p < .05$ level, ** $p < .01$ level, *** $p < .001$ level, (all two-tailed)

As shown in the Table, for the pink page, for example, correlations were highly significant for the shy ($r = .696$) and the non-shy ($r = .734$) groups, but weak,

albeit still significant, for the middle group ($r=.321$). In the case of the pastel purple and purple pages, the middle group correlations were not significant. Therefore, it is reasonable to conclude that participants who scored in the middle on the RCBS appear to view trust and visual appeal as more discrete concepts, when judgements are based on colour, than do the shy and non-shy groups.

Of interest was a finding that when all 840 scores for trust were correlated with the RCBS shyness scores, the parametric correlation was $r(838) = -.18$ ($p<.001$) with an explained variance of $R^2 = .03$.

Summary of Results

A summary of the results of Study 3 is shown in Table 19 below.

Table 19

Hypothesis number, prediction, test used and result

	Prediction	Test used	Result
H1	For a general population, varying the colours used in a home page would alter ratings of trustworthiness.	One-way ANOVA, matched-pairs t-tests, Friedman, Wilcoxon matched-pairs	supported.
H2	For a general population, varying the colours used in a home page would alter ratings of visual appeal.	One-way ANOVA, matched-pairs t-tests, Friedman, Wilcoxon matched-pairs	supported.
H3	The trust scores would be highly correlated with the visual appeal scores.	Pearson and Spearman correlations.	supported.
H4	Shy people would rate the pastel shaded pages more trustworthy than the same page with saturated color.	Matched-pairs t-tests, Wilcoxon matched-pairs	not supported.
H5	Shy people would rate the pastel shaded pages visually-appealing than the same page with saturated color.	Matched-pairs t-tests, Wilcoxon matched-pairs	not supported.
H6	Shy people would rate the pastel shaded pages more trustworthy than the non-shy people.	Independent samples t-test	not supported.
H7	Shy people would rate the pastel shaded pages more visually-appealing than the non-shy people.	Independent samples t-test	not supported.
H8	Shy people would rate the saturated-colour pages less trustworthy than the non-shy people.	Independent samples t-test	supported.
H9	Shy people would rate the saturated-colour pages less visually-appealing than the non-shy people.	Independent samples t-test	not supported.

Discussion

The colour of a simple Web page does apparently make a measurable difference to trustworthiness and visual appeal. The most interesting result was that the analysis of all participants' trust and visual appeal scores generated a series of Pearson correlations with a broad range. The present correlation of trustworthiness

and attractiveness would seem to confirm Ohanian's (1990) and Lindgaard et al.'s (2011) findings, but they go further, by showing that these correlations depend on the colour of the particular Web page and the group of people being tested.

However, the most decisive differences, those between the pastel yellow page (Figure 16 below) and the others, deserve a comment.

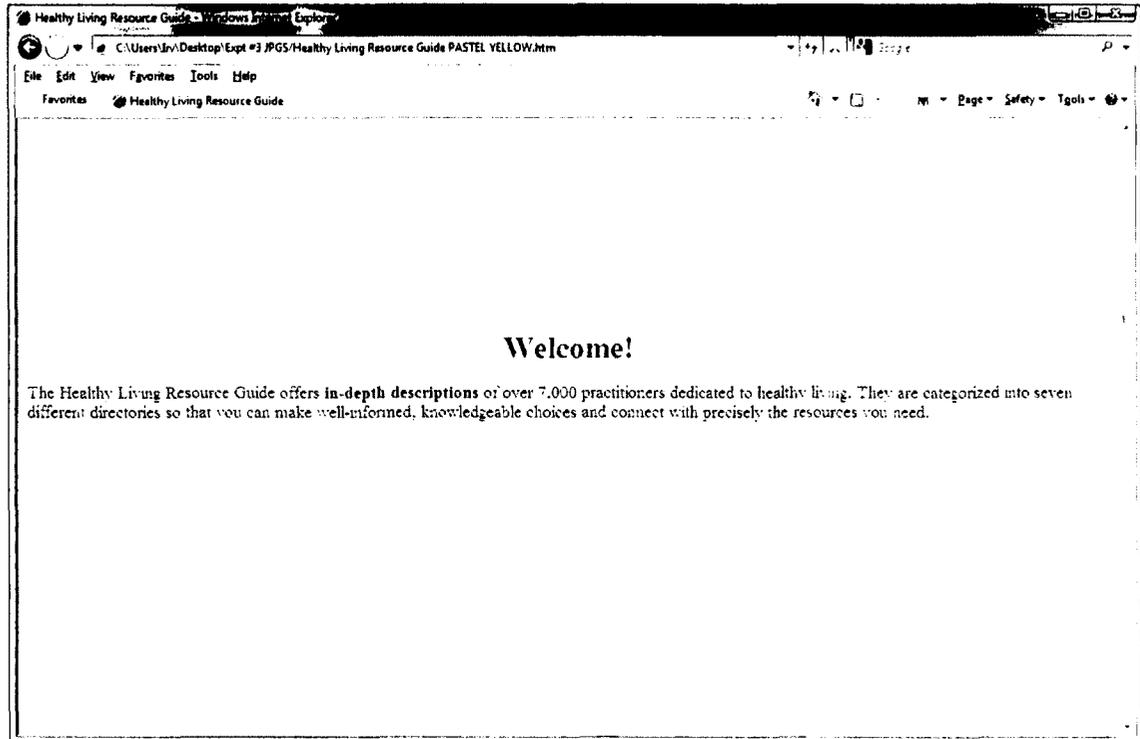


Figure 16. Pastel Yellow Page.

All graphics on the Web pages are in pastel yellow with the black text against a white background. One interesting, slightly disconcerting effect when looking at this page, somewhat akin to that experienced when looking at a Moiré image, in the sense that it is as if the eye tries to find a pattern but cannot grasp the pattern. The trust differences between blue and yellow, and blue and pastel purple could be related to this visually-dissatisfying experience. A questionnaire soliciting

comments concerning the subjective impressions of participants of the various pages might bring some verification of this explanation.

This study has shown that, for most colours tested, shy people rated Web pages the same as non-shy people. However, it is interesting that the shy group ranked the colour pink *less* trustworthy than the non-shy group. Anecdotally, one participant reported during debriefing to being very fond of the colour pink, but distrusting it as many Web sites used that colour to entrap the user. It is possible that other participants may have shared that view of pink as symbol for the fight against breast cancer, feeling it is being used cynically by companies that, in fact, often produce products that do contribute to cancer, such as cigarettes.

It was noted that, in general, the non-shy group exhibited a wider range of scores than the shy group. They were thus more likely to use the whole scale suggesting that the cautious shy person may be guarded even when assigning ratings on a trust scale.

This study has also shown the effectiveness of the Jian trust scale which was useful in demonstrating the complicated connection between shyness, visual appeal, and trustworthiness.

Study 4

The purpose of Study 4 was to select stimuli for Study 5 to examine the effect of photos on the trustworthiness and visual appeal of web pages varying in trustworthiness. First, it was decided to use only “friendly faces” since these tend to be the images seen on Web sites. As well, it was thought that to establish “trustworthy faces” might be more severe than friendly faces and bias against the

shy group of participants. These faces were to be of both genders and of couples, in order to control for any differences based on these variables.

Method

Participants

Twenty participants, 11 female and 9 male, were recruited from the same age group and source as before. None had participated in any of the previous studies. All were recompensed with 1% added to their Psychology course mark. No assessment of shyness was made. Each participant was tested individually in a session that lasted approximately 45 minutes.

Apparatus and Materials

The apparatus was the same as in Study 1 (p.80). The materials included informed consent forms, debriefing forms, and pens. Twenty-seven photos were copied from various internet pages, photos that, to the researcher, had the look of “friendly faces”. This was a large enough sample, that, if there were variations in the “friendliness” factor, the study would be able to identify two friendliest looking photographs from each group. The photos were found using Google and the search words “ healthy friendly faces”, “healthy friendly faces men”, “healthy friendly faces women”, “healthy friendly faces couples”. Nine photos were chosen for each group of “man alone”, “woman alone”, as well as “couples”. All appeared to be dressed in nurses’ garments. Most individuals were white, except for two African-American men. Of the couples, four were both white, with one of the men bearded; two were of a white woman and an African-American man; two were of an African-American woman and a white man; one was of an African-American man

with a white man (Appendix Y). It was thought that the use of a single ethnic type could be problematic. The photos were frontal head and shoulder shots.

An 8-item 7-point semantic-differential scale was constructed (Appendices Z to AC) to measure the friendliness of the photographs. The scale items were balanced left to right, with four of the items having the “agree” anchor to the left and four to the right, to control for any tendency to respond in a particular direction. Four items were of the form “This is a [.....] face; four of the form “This face is [.....]”.

Design

This study involved a repeated measures design with a single experimental procedure and 27 photos as the treatment levels. There was one dependent variable, the 8-item scale (Appendix AA), designed to measure “friendliness”. Both the presentation of the stimuli and the scale items, were randomized using the functions in MediaLab 2008 to prevent any order effect.

Procedure

To begin with, these participants were asked to sign a consent form (Appendix G). The 27 photos were shown one at a time, using a laptop computer. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants should look at the photo for a few seconds, then click a “Continue “ button, which would take them to a series of items, to be rated on a 7-item semantic differential scale. They were told there would be a smaller version of the photo accompanying the item and scale so memory was not an issue. After responding to an item, the next item would

appear, accompanied by the same small image. Afterwards, each participant was debriefed, thanked, and excused.

Results and Discussion

Each item was correlated with the others and with the total score for the 27 pages (Appendix AB). All the correlations were significant at the $p < .001$ level (2-tailed), indicating that the items were, at least partially, focused on a single concept, likely “friendly”.

Using the non-parametric Friedman analysis, the nine photos for couples, men, and women were ranked within their group, as shown in Table 20.

Table 20

Friedman rank order for the 3 photo sets

Photo of a couple	Mean Rank	Photo of a man	Mean Rank	Photo of a woman	Mean Rank
#9	6.20	#4	7.23	#5	6.43
#7	6.20	#3	6.33	#6	5.93
#1	5.88	#5	6.00	#3	5.75
#2	5.73	#9	5.53	#2	5.15
#4	5.20	#8	4.78	#9	5.10
#5	5.03	#6	4.68	#8	4.83
#6	4.70	#2	4.63	#1	4.73
#3	3.98	#1	2.95	#4	4.40
#8	2.10	#7	2.90	#7	2.70

Kendall’s W was calculated for each group, and yielded values of 0.247, .299 for the men, and .171 for the women (all $p < .001$), showing a significant amount of agreement among the participants. This was a useful statistic as it indicated that the results were based on consensus and not determined by a few of the participants, but that the ratings.

The highest ranking photos from each group were selected to be used in Study 5. Those photos are shown in Figure 17.

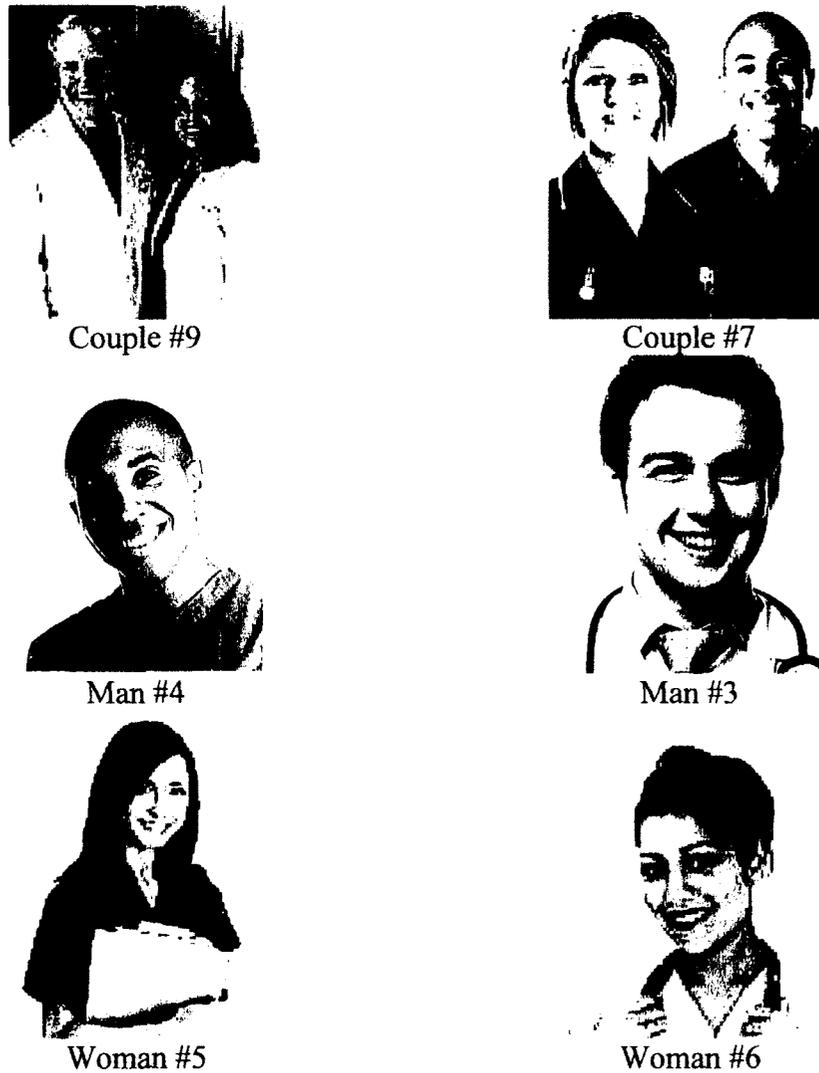


Figure 17. Six Photos used in Study 5

The goal of this study to identify some photos of “friendly faces” was accomplished. The two highest-ranking photos for each of the categories (“couples”, “female”, and “male”) were chosen (Appendix AD) to be placed on the Web pages in Study #5.

Study 5

The purpose of Study 5 to examine the effect of the photos of “friendly faces” on the trustworthiness and visual appeal of two home pages, one rated as very trustworthy, one as the least trustworthy in both Study 1 and Study 2.

Recall that Riegelsberger et al. (2003) found that Web sites originally perceived as credible without a photo, gained no credibility when a photo was added, but less credible sites did. Accordingly, Hypothesis 1 predicted that the addition of a photo to the less trusted Web page would yield higher trust ratings overall compared with the original page.

Along similar lines, Hypothesis 2 predicted that the addition of a photo to the most trusted Web page would yield lower trust ratings overall compared with the original page.

Since shy people tend to withdraw from situations that carry potentially negative affect (Crozier, 2005), they are more likely to avoid re-using of a site that presents a poor initial experience. A Web site with a photo of a person may indicate potential social interaction, which could lead the shy person to interpret this negatively and experience fear and thus avoid spending time on the site. Therefore, Hypothesis 3 predicted that participants in the shy group would give lower trust ratings for a Web page with a photo image of people than for the same page with no added image.

Finally, a subset of users identified by Riegelsberger and Sasse (2002) were very risk-aware, did not trust on-line vendors, and saw photos as manipulative. This

distrustful group could very well include some or all of the shy people in the present study. Thus, Hypothesis 4 predicted that participants in the shy group would trust the photo-altered pages less than the non-shy group.

Lindgaard et al. (2011) have demonstrated that the trustworthiness of a Web site is connected to the visual appeal aspect of a Web site. Study 3 in this thesis supported a relationship between trust and visual appeal. Hypothesis 5 therefore predicted that the trust scores would be highly correlated with the visual appeal scores.

Table 21 presents the five hypotheses.

Table 21

Five Hypotheses for Study 5

Hypothesis 1	The addition of a photo to the <u>less</u> trusted Web page would yield higher trust ratings overall compared with the original page.
Hypothesis 2	The addition of a photo to the <u>most</u> trusted Web page would yield lower trust ratings overall compared with the original page.
Hypothesis 3	Participants in the shy group would give lower trust ratings for a Web page with a photo image of people than for the same page with no added image.
Hypothesis 4	Participants in the shy group would trust the photo-altered pages less than the non-shy group.
Hypothesis 5	The trust scores would be highly correlated with the visual appeal scores.

Method

Participants

Sixty-one participants, 31 female and 30 male, were recruited from the same age group and source as before. All of the undergraduates were recompensed with 1% added to their Psychology course mark. Each participant was tested individually in a session that lasted approximately 45 minutes.

Sample Size

Since Hypotheses 3 and 4 dealt with evaluations by shy and non-shy participants, I required a sufficiently large sample to select 20 shy and 20 non-shy participants for these analyses. In 2007, 1038 Carleton University Introductory Psychology students who filled out the RCBS scale (Cheek, 1983) demonstrated a range of scores between 13 and 62, with 1/3 below 29 and 1/3 above 40. I decided that, if I tested 60 participants, I could choose the highest and lowest one-thirds on the RCBS for the two groups.

Apparatus and Materials

The apparatus was the same as in Study 1 (p.80). The materials included informed consent forms, debriefing forms, and pens. Two Web pages were chosen for this study, the Manitoba government site, rated highly on trust in Studies 1 and 2, and the ArticleWorld site, given the lowest trust ratings in Studies 1 and 2. Seven forms of each page were generated using KompoZer v. 0.7.10 (<http://www.kompozer.net/>): the original page with no face (i.e. unaltered, as in Study 2), two with a male friendly face, two with a female friendly face, and two with a couple with friendly faces. This resulted in fourteen Web pages (Appendix AE).

The photos were placed to the left side of the Web pages, which is a salient part of the scanning zone for people browsing the Web. The images were approximately 170 pixels wide by 200 pixels tall. Examples of the Web pages are shown in Figure 18.

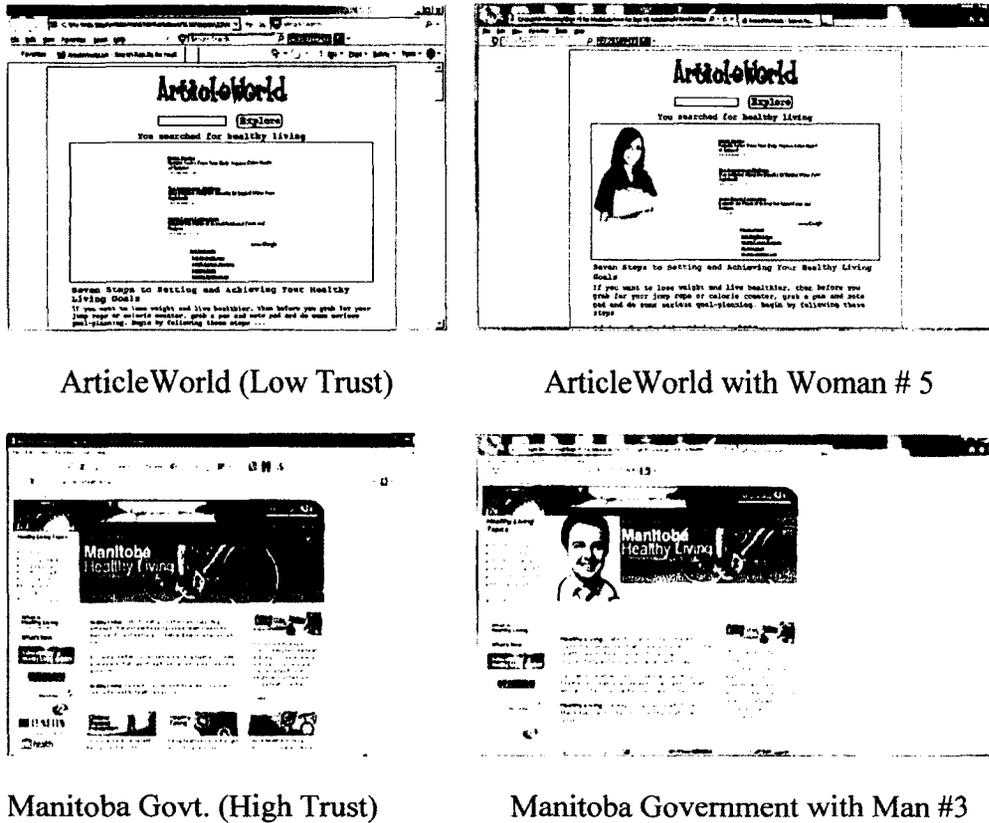


Figure 18. Low- and High-Trust Web Pages with and without Photos.

Each participant completed out a 20-item scale (Appendix Q) for each of the 14 Web pages. This scale comprised the 11-item Trust scale tested in Study 2 and the 9-item Lavie and Tractinsky Visual Appeal scale. Each item was scored on a 7-point semantic differential scale. As well, the participants filled out the 13-item RCBS Shyness scale as in Study 3.

Design

This study employed a split plot factorial design in a 1x14 (1 experimental procedure, 14 Web pages) repeated measures design.

The design included two within-subjects independent variables, photos, with seven levels (plain, two male, two female, two couples) and Web page type with two levels (high-trust, low-trust); and one between-subjects independent variable,

shyness, with two levels (shy, non-shy). Three dependent measures were recorded, Trust, Visual Appeal, and Shyness. The presentation of the stimuli and the scale items was randomized using MediaLab 2008 v. 1.33 to control for order effects. The Trust and Visual appeal items were not separated in the administration of the items. The Agree/Disagree anchors were reversed for 11 of the 20 items to control for any tendency to respond in a particular direction.

Procedure

Upon signing a consent form (Appendix G), each participant viewed the 14 home page images, one at a time, and rated each on the 20-item trust/visual appeal scale. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants were told that they could look at the Web page for a few seconds, then click a “Continue” button, which would take them to a series of statement, to be answered on a 7-item semantic differential scale. They were told there would be a smaller version of the Web page accompanying the items so memory was not an issue. After responding to each item, the next item would appear, accompanied by the small Web page image. At the end of the session, each participant filled out the 13-item RCBS Shyness scale, and was then debriefed and excused.

Results

The results are divided into six sections. First, (1) data exploration, followed by (2) the comparisons between pages based on the less trustworthy page, and (3) the comparisons between pages based on the more trustworthy page. Next (4) comparisons measuring the effect of a photo on shy participants, (5) comparisons

between the effect of a photo on the shy and non-shy participants, and, finally (6) correlations between trust and visual appeal scores.

Data Exploration

First, the distributions of responses (n=61) for the variables for the 14 Web pages, the Jian scale as well as the Visual Appeal scale, were examined, attending especially to skewness and kurtosis values with respect to their standard errors.

Skewness values around twice the value of the standard error indicate a marked deviation from normalcy. Four of the 14 trust variables and two of the 14 visual appeal variables displayed such ratios. As a result, I analyzed Study 5 non-parametrically.

Comparisons of Pages based on the Less Trustworthy Page

To test Hypothesis 1, that the presence of a photo would yield higher trust ratings compared with the less trusted non-photo page, the Friedman test was applied to the Jian trust scores for the seven ArticleWorld pages, the less trustworthy page. This generated a $\chi^2(6) = 14.248$ (p=.027), showing that the Web pages exhibited significant differences. Since the analysis is non-parametric, the medians and ranks for the pages are shown in Table 22.

Table 22

Medians and ranks for ArticleWorld, the Least-Trustworthy Page.

ArticleWorld pages	Median Trust score	Mean Friedman rank
Woman #5	31	4.53
Couple #9	32	4.47
Man#4	33	4.11
Couple #7	31	3.92
Woman #6	34	3.84
Man#3	29	3.78
Plain	30	3.34

The table shows that the median scores for the pages are very similar, ranging from 29 to 34 out of a possible 77 maximum score. I had decided to test the highest ranking photo-altered Web page against the plain ArticleWorld page, to increase the power of the analysis. Using the Wilcoxon Rank Sum tests (Appendix AF), one of the ArticleWorld pages with a female face (AW#5) was rated more trustworthy than the plain page, ($Z = 2.380$, $p = .017$). The two pages are shown in Figure 18 above. As well, I tested the second-highest rated site, one with a couple, AC#9, generating a $Z = 2.290$ ($p = .022$). Halving the overall significance level, $\alpha = .05$, this difference was also significant, even at a post hoc level. As an indication of the effect these differences, 39 of the 61 participants (64%) rated the *page with the image of a woman* (AW#5) as more trustworthy than the plain page. 37 (61%) rated the *page with the image of a couple* (AC#9) as more trustworthy than the plain page. These results supported Hypothesis 1, that the presence of a photo would yield higher trust ratings compared with the less trusted non-photo page.

As a matter of interest, I continued applying the Wilcoxon tests (Appendix AF) to the other four Web pages. These comparisons did not approach significance, suggesting that simply adding any photo would not alter the trust ratings. The photo in AW#5 was the highest ranking photo of a woman in Study 4.

Comparisons of Pages based on the More Trustworthy Page

To test Hypothesis 2, that the presence of a photo would yield lower trust ratings compared with the most trusted non-photo page, the Friedman test was applied to the Jian trust scores for the seven Manitoba pages, the most trustworthy page. This generated a $\chi^2(6) = 5.803$ ($p = .446$), showing that the Web pages

exhibited no significant differences suggesting that the addition of a photo had not diminished the trustworthiness of the original page. The medians and ranks for the pages are shown in Table 23.

Table 23

Median Trust scores and ranks for Manitoba, the Most-Trustworthy Page.

Manitoba pages	Median Trust score	Mean Friedman rank
Plain	64	4.34
Woman #5	60	4.17
Couple #9	61	4.11
Man#4	61	4.07
Woman #6	61	3.95
Couple #7	60	3.77
Man#3	60	3.57

Table 16 shows that the median scores for the pages were very similar, ranging from 60 to 64 out of a possible 77 maximum score. Even though the Friedman test was not significant, there is evidence (Zimmerman & Zumbo, 1993) that the Friedman test is not as powerful as the Wilcoxon matched-pairs signed-ranks test, so I applied the Wilcoxon test comparing the plain Manitoba page to the lowest-ranking Manitoba page with an image (MM#3), generating $Z=2.414$ ($p=.016$) suggesting that the addition of the photo had brought about a decrease in trustworthiness for the original Manitoba site.

This supported Hypothesis 2, that the presence of a photo would yield lower trust ratings compared with the most trusted non-photo page. Thirty-six of the 61 participants (59%) trusted the plain Manitoba page more than with the male face. As a matter of interest, I continued applying the Wilcoxon tests (Appendix AF) to the other five Web pages. These comparisons did not approach significance.

A particular photo thus did lessen the trustworthiness of the page, but it is not clear why Man #3 did so. The photo in MM#3 was the second highest ranking photo of a man in Study 4.

The Effect of a Photo on Shy Participants

Next, Hypothesis 3, that a face on a Web page would *decrease* trust for the shy group was tested. Two groups of 20 participants were selected using the RCBS scale. Participants with scores between 36 and 52 (median = 41.5) made up the shy group. Those with scores between 13 and 27 (median 22.0) were assigned to the non-shy group. There were 10 males and 10 females in the shy group, 13 males and 7 females in the non-shy group. The Mann-Whitney *U* statistic (Mann and Whitney, 1947) showed these two sets of scores to be significantly different ($U = .000$, $Z = -5.425$, $p < .001$). It was reasonable to use the two groups of shy and non-shy people.

To examine this, I applied the Friedman test to the trust scores *within the shy group*. This generated a $\chi^2(6) = 10.879$ ($p = .092$) for the ArticleWorld sites, $\chi^2(6) = 1.293$, $p = .972$) for the Manitoba sites, suggesting that neither set of pages exhibited significant differences. To test this further, the Wilcoxon matched-pairs signed ranks test found the plain Manitoba page compared to the lowest-ranking modified page, with Man #3, demonstrated a non-significant $Z = 0.396$, $p = .163$. Similarly, the plain ArticleWorld page compared to the lowest ranking modified page, with Couple #9) demonstrated a $Z = 1.767$, $p = .077$. Hypothesis 3 was therefore not supported.

The Effect of a Photo on Shy and Non-Shy Participants

Hypothesis 4, predicting that the shy group would trust the various photo-altered pages less than the non-shy group, tested by the Mann-Whitney U test, showed no significant differences. Hypothesis 4 was therefore not supported.

An examination of the mean ranks (Appendix AG) and the median trust scores (Appendix AH) for the various pages between shy and non-shy groups found an interesting pattern. The trust ratings for the ArticleWorld (AW) sites were such that the shy group found all of the ArticleWorld sites somewhat more trustworthy than the non-shy group. Conversely, the non-shy group found all of the Manitoba sites somewhat more trustworthy than the shy group.

The non-shy group found the more trustworthy pages (Manitoba) more trustworthy than the shy group, and the untrustworthy pages (ArticleWorld) less trustworthy than the shy group. The non-shy group assigned more extreme scores than the shy group. The shy group tended to use the middle of the 7-point scale for their ratings, yielding lower scores for the more trustworthy pages, and higher scores for the less trustworthy pages. The same pattern was found for the visual appeal scores except for two pages, Shy AM#4 (median = 23) and non-shy AW#6 (median = 23.5).

This initiated analysis of the 140 scores for these two groups (20 participants and 14 pages). Again, these data sets were highly skewed. Because of the extreme skewness values, the 140 scores from each group were subjected to the non-parametric Mann-Whitney two-sample rank sum test, ranking the two groups within the two sets of pages, that is shy/Manitoba versus non-shy/Manitoba, and

shy/ArticleWorld versus non-shy/ArticleWorld. The Mann-Whitney results are shown below Table 24.

Table 24. Mann-Whitney Results.

	Shy median	Non-Shy median	U Statistic
Manitoba sites			
Trust	59	66	7349.0 ***
Visual Appeal	40	43	7738.5 **
ArticleWorld sites			
Trust	32	27	7739.0 **
Visual Appeal	29	22	6468.5***

p<.01, *p<.001(both 2-tailed)

As the Table shows, the shy group differed significantly from the non-shy in the comparisons of both trust and visual appeal scores.

Correlations of Trust and Visual Appeal

Next I tested Hypothesis 5, that the trust and visual appeal scores would be highly correlated. Looking at the entire sample of 61 participants, the Jian trust scores were correlated with the Visual Appeal scores for (1) all of the 14 pages, (2) for the seven ArticleWorld pages as a group, (3) for the seven Manitoba pages as a group, and (4) for each page individually (Table 25). All correlations were calculated using the Spearman rank correlation coefficient.

For all of the 14 pages, the correlation was highly significant($\rho = .819$, $p < .001$), as it was for the seven ArticleWorld pages as a group ($\rho = .703$, $p < .001$), and for the seven Manitoba pages as a group ($\rho = .570$, $p < .001$)

Table 25

Web page	Spearman Correlation
Plain Manitoba	.540***
Manitoba with Couple #7	.577***
Manitoba with Couple #9	.579***
Manitoba with Man #3	.562***
Manitoba with Man #4	.610***
Manitoba with Woman #5	.552***
Manitoba with Woman #6	.543***
Plain ArticleWorld	.643***
ArticleWorld with Couple #7	.741***
ArticleWorld with Couple #9	.745***
ArticleWorld with Man #3	.716***
ArticleWorld with Man #4	.700***
ArticleWorld with Woman #5	.701***
ArticleWorld with Woman #6	.659***

*** significant at $p < .001$ level (2-tailed)

As Table 25 shows, comparisons for each of the Web pages were all highly significant. The lowest correlation ($\rho(59) = .540$), was for the plain Manitoba page "MP". The highest, ($\rho(59) = .745$), was for AC#9, the ArticleWorld page with a couple. All of the ArticleWorld pages had higher Trust/ Visual Appeal correlations than the Manitoba pages. Looking at the scatterplot in Figure 19, the ArticleWorld pages are clustered relatively closely around low trust and visual appeal scores, while the Manitoba pages show higher trust and visual appeal scores, but are more scattered.

ArticleWorld sites (Blue) and Manitoba sites(Yellow)

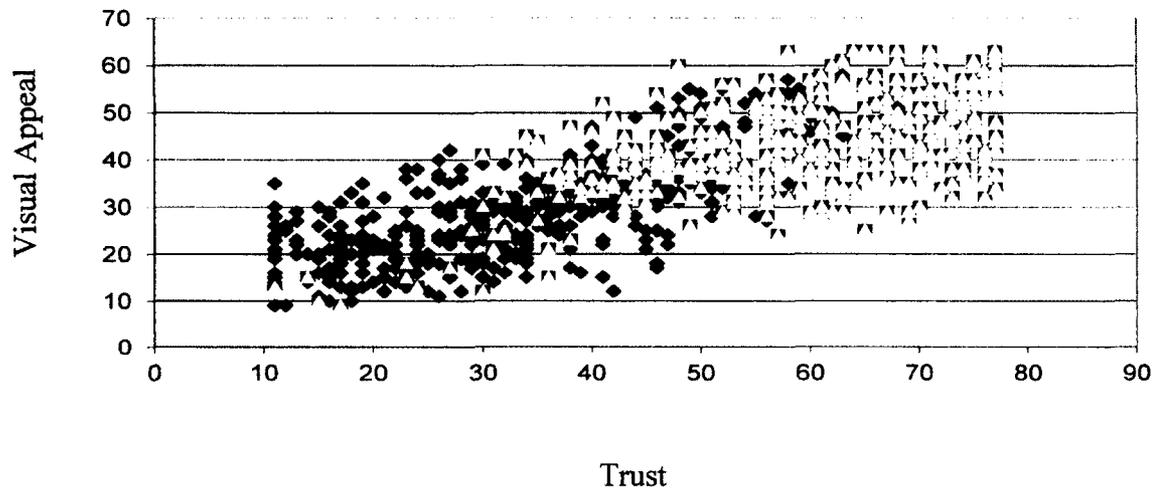


Figure 19. Trust/Visual Appeal scores (n=854)

The plain Manitoba site was the most trustworthy site originally. However, the correlation of trust and visual appeal was weaker, albeit still highly significant ($p < .001$), than for any less trustworthy site. The higher correlation for all 14 sites together is due to the marked difference between the ratings for the Manitoba sites and the ArticleWorld sites. Thus, the relationship between trust and visual appeal varies, depending on the source of the data.

A home page with lower trustworthiness ratings resulted in a higher correlation of visual appeal and trust.

Of interest was a finding that when all 854 scores for trust were correlated with the RCBS shyness scores, the parametric correlation was $r(852) = -.05$, a non-significant statistic. ($p = .16$). Overall, there was no relationship between shyness and trust in these web pages.

Summary of Hypotheses

Table 26

Hypothesis number, prediction, test used and result

	Prediction	Test used	Result
H1	The presence of a photo would yield higher trust ratings compared with the <u>less</u> trusted non-photo page.	Friedman and Wilcoxon	Partially Supported
H2	The presence of a photo would yield lower trust ratings compared with the <u>most</u> trusted non-photo page.	Friedman and Wilcoxon	Partially Supported
H3	The shy group would give lower trust ratings for a Web page with a photo image of people than for the same page with no added image.	Mann-Whitney	Not Supported
H4	The shy group would trust the various photo-altered pages less than the non-shy group, i.e. lower trust ratings would be assigned by the shy than the non-shy participants.	Mann-Whitney	Not Supported
H5	The trust scores would be highly correlated with the visual appeal scores.	Spearman correlations	Partially supported

Discussion

This study supported earlier research by showing that a Web page that was perceived originally as trustworthy without a photo, gained no trustworthiness when a photo was added, and, in fact, lost trustworthiness with the addition of one of the photos. The less trustworthy page gained from the addition of two of the photos.

The most interesting finding was that the correlation between the trust scores and the visual appeal scores depended very much on the data set addressed. For the

total sample of participants, the trust and visual appeal scores yielded a higher correlation, when all pages correlated, than when the seven-page groups were correlated. This was due to the underlying difference in trustworthiness of the two sets of pages. This finding adds to the research on trust and visual appeal. At first glance, one could say that visual appeal must account for a large factor in initial trust, mirroring the literature on visual appeal and trust judgements in face-to-face situations (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992; Langlois et al., 2000). But the degree to which visual appeal correlated with trust varied by the rated trustworthiness of each page, possibly due to the variety of factors contributing to initial trust in a Web page. The less trustworthy a page was without an added image, the more trust was correlated with visual appeal. In that case, the only prominent variable for the user was the visual appeal of the page, resulting in very high correlations. Conversely, the more trustworthy page may have had more design factors contributing to trust, such as colour, use of graphics, reputable name, clean layout, all identified by Sillence et al. (2004, 2007) and Fogg, Soohoo et al. (2002) as supporting initial trust. These factors were sufficiently salient to weaken the effect of visual appeal on trust.

A second finding had to do with the shy group tending to group their responses toward the center on the 7-point scale. This is what Guilford (1954) called “the error of central tendency. This may be of some interest in the application of the RCBS, since shy people who fill out that scale may avoid the extremes on the various items, and therefore, rate themselves as less shy than they actually are.

No differences were found for shy people versus non-shy people in their evaluations of trustworthiness of a Web page based on the addition of a photo of a human face.

General Discussion

This chapter summarizes the main findings of the five studies detailed in this thesis and presents the contributions these findings make to the overall study of the influence of colour, photos, and shyness on trust and visual appeal in Web site design. I follow this with some directions for future research and a conclusion.

Summary of the Main Findings and Contributions

The studies in this thesis were designed to explore how trust and visual appeal of a Web page can be modified by the use of colour and photos. As well, they examined potential differences between shy and non-shy people.

The first two studies validated a trust scale to measure affect-based initial trust in a Web page. The items on the trust scale had been shown by Jian et al (2000) to tap into trust, but the Web sites used as stimuli were selected by scale items based on design look. Thus Study 2 lent support to the connection between design look and trust.

The significant positive correlations in Studies 1 and 2 support the view that the researchers were tapping into a single construct, initial trust in a Web page. This construct appeared to be multi-faceted, as indicated by the variety of phrases in the scales: design layout, navigation, colour and graphics, as well as the affective dimensions on which the participants attributed characteristics to the Web page such as deception, reliability, and concern for the user.

It was interesting to see how easily participants assigned home pages different scores with respect to the affective items in the trust scales. Trust in a person may include such terms as “deceptive” and “has integrity”; it is less obvious how these can describe a Web page, especially on first sight. Trust in an object, a car say, normally would be based on perceived expertise, past experience with the vehicle or its brand, or derived trust as when a friend gives a recommendation. These sources were not available here, yet this did not cause the participants to waver or to randomly assign ratings. These undergraduate students had no problem with the idea of trusting or not trusting a Web page. Perhaps their experience with internet security concerns has created the need to make such an evaluation at once, when looking at a Web page, just as one makes such decisions when walking in an unknown neighbourhood of a city.

Study 3 examined the effect of colour of a Web page on trust and visual appeal rating for a general population as well as for shy and non-shy groups. The study was able to show some effects for colour, as well as suggesting that shy people use the rating scale differently from a non-shy group by exhibiting a more-conservative use of the range of options.

Three colours are worthy of mention, when looking at the relation between the ratings for trustworthiness and visual appeal. Blue was shown to be a good colour choice for Web design as it rated first in both categories. This supported traditional design directions, as many large, successful companies, such as IBM, Bell, and CareFirst BlueCross, use this colour in their logos and Web sites. It is possible there was a halo effect (Nisbett & Wilson, 1977b) derived from the

existing, well-trusted brands, associated with this colour. The initial, early-Internet choice of blue by these companies may have been based on studies showing a marked preference for blue in the colour preference studies referenced in the literature review. This may be why Jian and his team kept the word “familiar” in the trust scale despite the fact that their participants rated it as “not related to trust”. Even though this thesis dealt with initial evaluation of a Web page, the colour blue is so ubiquitous that likely it was familiar to the participants, and therefore perceived as more trustworthy. The inherent visual appeal of blue has been well-documented, and so may explain its position as number one in both categories. This fact also supported the decision not to use the word “familiar” in the adapted trust scale, as this would have tended to favour familiar colours, and disadvantage less common ones.

Pastel blue is often used in Web site design today. It ranked fifth of the 10 colours on trustworthiness here, but second with respect to visual appeal. It is possible that the traditional Web designers may be equating colour preference, visual appeal, with trustworthiness. However, they may not have studied trustworthiness. This thesis contributes to this area of Web design.

Study 3 results showed that the colour of a simple home page made a significant difference in trustworthiness for the participants who rated the pastel yellow page less trustworthy and the blue page was rated more trustworthy than some of the others.

Using stimuli identified in Study 4, Study 5 examined the effect of a photo on a Web page on trust and visual appeal for a general population as well as for shy

and non-shy groups. Study 5 confirmed that some photos yielded higher trust ratings over the least trusted non-photo page and lower trust ratings over the most trusted non-photo page. This study showed that shy people tended to use the middle points on the rating scales.

The most interesting finding in Study 5 was that the correlation between the trust scores and the visual appeal scores depended very much on the data set addressed. For the total sample of participants, the trust and visual appeal scores yielded a higher correlation when all pages correlated than when the seven-page groups were correlated. This was due to the underlying difference in trustworthiness of the two sets of pages. This finding adds to the research on trust and visual appeal.

The use in this thesis of the Wilcoxon results to show the number of participants who actually trusted one page more than another, or preferred one page to another visually, demonstrated a tool which may provide more accurate details directly applicable to the concrete decisions of Web site designers.

Theoretical Implications

This thesis has drawn on three bodies of research; visual appeal and Web design, visual appeal and trust, and shyness and trust, and has made a contribution to each area.

First, this thesis has shown that some of the comments made by the participants in Sillence et al. (2004, 2007) and Fogg et al. (2000) concerning Web design are valid indicators of initial trust in a Web page. Sillence et al. (2004, 2007) and Fogg et al. (2002) proposed initial trust in a Web site primarily dependent on

design. It was phrases from their qualitative research that formed the basis for the trust scale used in Study 1. Lindgaard et al. (2011) found that this initial trust stage was based on visual design characteristics such as ‘boring web design’ and ‘use of colour’, descriptors similar to those found by Lindgaard, et al (2006) to be highly correlated with visual appeal, namely ‘interesting/boring’, and ‘good/bad colour’. This thesis has shown that some design look descriptors are related to trust. In Study 2, it showed that those design look/ trust descriptors correlated with both the McKnight and the Jian trust scales. In Studies 3 and 5, it was shown that trust and visual appeal are related. These connections support the work of Sillence et al. (2004, 2006, 2007), and Lindgaard et al. (2006, 2011) of the relationship between design look, visual appeal, and trust. It has quantified the relationship of visual design and trust and visual appeal and trust and shown, in Study 5, that the connection is stronger for less trustworthy sites and weaker for more trustworthy sites. This thesis has also shown that the assessment of design look factors, such as ‘corporate look and feel’, can be related to trust in entirely different ways, depending on the participants’ demographics, middle-aged women or twenty year old students. As well, the women had a vested health-oriented interest in the Web sites, whereas the Web pages in this thesis had no such personal connection to the students.

There is evidence that visual appeal evaluations carry over into other judgements (Thorndike, 1920; Nisbett & Wilson, 1977) including web trustworthiness ratings (Lindgaard et al., 2011). This halo effect has been supported by this thesis in that there were significant correlations between visual appeal and

trust scores. However, the correlations in Study 5 showed that the entanglement of these two constructs is higher when viewing a less complex and less trustworthy web page than for one that was more complex and more trustworthy. Also, it was seen that the correlations were higher for both shy and non-shy participants, and less so for those who were rated in the middle on the shyness scale.

This thesis found blue and pastel blue were more visually-appealing than pastel purple. Sillence et al. (2004) mention “pale blue”, and “insipid green backdrop”, as examples of design issues that could lead to mistrust of a Web page. Study 3 in this thesis did not find strong distrust or visual aversion to those colours, although the colours here were of graphics, not text or background, which, may have been the aspects referred to by Sillence et al.’s participants. This makes it clear that when colour and Web pages are discussed, the text, background, and graphics colours should be noted, as it is in the interaction of these three that visual appeal and, perhaps, trust decisions are made. As well, these comments may come from the tendency of people to find justification for their decisions, even if this justification has no basis in fact. The Sillence and Fogg participants mistrusted the sites, perhaps, not due to those particular colours, but they still reported them as the reasons. This thesis actually measured the reactions of participants to these colours. This strictly behavioural measure may be more informative than the comments.

This thesis supported the work of Fogg and Tseng (1999), in that modifications in on-screen characters, specifically graphics and images did affect trustworthiness and visual appeal as they predicted.

Second, this thesis has added some validation to the use of the Jian Trust in a Machine Scale especially as used in evaluating initial trust in a Web page. The view that trust and distrust are unidimensional as presented by Gambetta (2000) and Jian (1998) was supported in Study 2 since the correlations between the “I can trust this Web Page” item, and the five distrust-oriented items were all significantly negative and those with the five trust-oriented items were all significantly positive. This did not support the two-dimensional view of Lewicki et al (1998). Their recent view of trust and distrust seems to be based on a misunderstanding of facet theory and the actual conclusions of certain studies. The utility of the two-dimensional view of trust and distrust would seem to be better served by the trust + expertise = credibility model of Fogg et al (2000), and careful separation of co-existing trust/distrust focusses such as in Clarke and Payne (1997).

Third, it has shown that certain colours are not simply preferred and assigned higher visual appeal values, but are also more or less trustworthy. This supports Fogg et al.'s (2003) prominence theory by showing that when a Web page is not trusted, visual appeal becomes more prominent, and the correlation between visual appeal and trust is increased over a more trustworthy page.

This research has supported previous findings that blue is given higher visual appeal ratings, than other colours. It has supported Cyr, Head, and Larios (2010) who found that a group of Canadian participants group rated a blue web page significantly more appealing and trustworthy than a comparable yellow page. As well, Cyr et al. found that yellow was significantly disliked. This thesis found the yellow page to be placed second last for both visual appeal and trust. Culture and

demographics are important variables in the assessments of colour and visual appeal.

The very low ratings given for the pastel yellow graphics against a white background are reminiscent of the findings of Noiwan and Norcio (2006) whose American participants rated both the white text on yellow and black text on yellow banners lowest for four usability categories. Only 3% of that group rated yellow as a preferred colour. The usability issue was possibly confounded in this thesis with visual appeal, and future research should involve a test of usability perception.

Shy people are thought to be anxious and earlier literature found that high-anxiety people prefer pastel colours over more saturated shades. As well, there was evidence that shy people with low arousal appeared to find faces more prominent, and I had expected to find that this would cause aversion and distrust of a Web page with an added photo. This thesis did not find that shyness affected judgments of trustworthiness or visual appeal of colours or of Web pages with added photos. However, the sample size is the likely source of the lack of support for the hypotheses. The shy group in Pozzulo et al. (2005) was the shyest 2.5% of 841 participants, with RCBS scores above 44. The shy groups in the studies in this thesis were drawn from the top 23% and 33% of much smaller sample, with shyness scores above 41 and 36 respectively. The level of shyness of these participants may not have been high enough to generate the expected results.

Relevant to the issue of sample size for the shy and non-shy groups, correlations of the various trust and visual appeal variables with the RCBS scores in Studies 3 and 5 found only small correlations for the blue, pink, pastel yellow, man

#3 and man #4 pages. Subsequent analyses showed no advantage to using the complete samples in either study. Further research with larger samples may find greater discrimination.

The small sample size does not explain why this thesis did not support the Kim and Moon (1998) findings that cool, pastel tones, pastel background, and low brightness stimulated more trust than bright, primary colours and background. Noiwan and Norcio (2006), in explaining differences between the American and Thai participants speculated that the Thai group was more anxious than the Americans. As well, 30% of the Thai group rated yellow as their 1st or 2nd preferred colour, only 5% of the Americans did so. The participants in Kim and Moon (1998) were South Korean, and they mention the differences in results due to diverse cultures. As well, their interface was for cyber-banking, and could elicit different reactions than a health-related page.

Fourth, this thesis has supported earlier research showing that the addition of photos to a Web page can have differing effects. It has shown that the addition of the friendliest male face to a trustworthy Web page actually diminished its trustworthiness to a significant degree. Other studies have used photos rated for trustworthiness and identified as “Customer Care” employees, for on-line shopping sites, and generic photos embedded in cyber-banking sites. Riegelsberger et al. (2003) allowed the participants to browse the site and access information. These were different scenarios than in this thesis which used friendly faces on ‘healthy living’ pages with no role identifier. It is interesting that the photo that affected the trust ratings for the more trustworthy pages (Manitoba Government) was the

friendliest male face with no role identifier. It is possible that a very friendly face on a health-oriented Web page may be very prominent, and inappropriately over-friendly for the site, and therefore seen as a commercialization, and thereby reduce trust.

This thesis has shown that the interplay between colour, photos, individual web page, trustworthiness and shyness is real but quite variable. Generalizing from this work to other genres of sites, or other groups of people would be inappropriate. However, this thesis has shown that there is reason to continue to study these constructs especially if a web designer has a defined clientele. Many of the hypotheses were not confirmed because the variability amongst the participants was great enough to overshadow any variability between the shy and non-shy groups on the questions addressed here.

Future Research

The results in Study 3 suggested that Web designers should avoid pastel yellow, yellow, even pastel purple, and instead choose blue to inspire trust and visual appeal, but the other colours appeared to be neutral. It would be interesting to use a variety of Web pages, some using primarily a single colour, others with a few colours, to see how these would rate for trust.

This thesis was designed to evaluate initial trust in a Web page, but no check was made for the possibility that the participants may have visited these sites earlier. This was especially true for the WeightWatchers site which is relatively well-known. A future study might make use of sites that are familiar and some that are not, perhaps constructed specifically for the study.

A replication of this study should also use the Ishihara test for colour blindness a condition that affects about 7 % of males, but only 0.5 % of females. That was not done here.

Several comments were made spontaneously which may be valuable to note, since these appeared to be cogent explanations. One participant said that the “government” aspect of the Manitoba site overrode all other considerations and for that reason, for that person, that was the more trusted site. A second, equally spontaneous, comment was that governments were trying to manipulate the population concerning health and therefore, for that participant, that site was untrustworthy. Two sophisticated computer users, computer science students, were distrusting of the photo-embedded sites, because, they said, photos were a locus for “trojans” and other hacking techniques. These reasons indicate the degree of sophisticated thought, even in an initial evaluation of a Web page, that can be at the disposal of a more-fluent computer user. Future research could test for attitudes to specific site sponsors, such as governments, and computer experience, to see how these variables affected trust and visual appeal ratings.

Study 3 used one Web page, altered by colour, so as to control for other design look features such as navigation cues. A future study could use a between-subjects design, with 100 Web pages, of various colours, but with each participant only viewing a subset, to allow for the time to complete the Trust/Visual Appeal scale. As well, the use of a 7-point semantic differential scale could be improved upon by the use of a continuous visual analogue scale (Reips and Funke, 2008),

which would generate interval data that would be more susceptible to parametric analysis.

The pages used in Study 3 were of variously coloured graphics against a white background. The contrast, or lack of such, between the graphics colour and the background likely confounded the results, especially as concerns the pastel yellow page. Colour is always seen in a context, and the pastel yellow graphic against white would likely elicit a different reaction from the same pastel yellow against pastel blue, or blue. Future research could study a large number of web pages with many combinations of graphics and background colour to specify that a particular graphic colour may be more trustworthy and visually appealing against certain backgrounds but not others.

There are many opportunities for extending the results of Study 5. The use of more trustworthy or less trustworthy faces, attractive and unattractive faces, or formal and casual, in different types of sites besides the health context used here would extend the research. This study could not examine this wider range of stimuli as the number of treatment levels was an issue.

A study involving different genres of home pages would be valuable to see how trust would vary in an initial evaluation of financial, entertainment or news sites. Colour would likely affect ratings differently for sites focussed on travel, sports, computer supplies or entertainment.

This thesis avoided using the “familiar” item from the Jian scale, to avoid any specific bias for the WeightWatchers site, or any other that might have been used by a participant before the studies. This may also have weighted the results for the

Manitoba Govt. site as might the very fact that it is a government site. This would not affect the fact that it was the most trusted site, and so does not weaken the results of this that study, or of the rankings in Studies 1 and 2. However, future research should question the participants with regards to previous interaction with the sites so as to provide evidence as to how powerful a predictor of trust this might be.

As well, it should be noted that Pozzulo et al. (2005) found differences in the attention to faces between shy and non-shy only when high anxiety obtained. There was no high anxiety in this laboratory scenario. Further research could manipulate that variable.

It is possible that there exist different subgroups of participants such as described by Riegelsberger and Sasse (2003) and further research could use the portions of the McKnight scale (McKnight et al., 2004) focussed on disposition to trust to further separate the shy and non-shy into two more groups, high trust and low trust, perhaps with some interaction effects.

Conclusion

This thesis aimed to demonstrate measurable differences in the perceived trustworthiness and visual appeal of Web sites varying in the colour of graphics, as well adding photos of “friendly faces”. It also aimed to show that the impact of such amendments may vary according to the test population, specifically shy and non-shy people. In particular, it showed that Web pages can be differentiated by trust ratings based on design look. It validated the Jian trust scale to measure affect-based initial trust in a Web page, and it demonstrated the connection between

design look and trust. In addition, the thesis showed clearly that shy people tended to use the rating scale more conservatively than the non-shy group. Finally, the thesis confirmed that the effect of photos of friendly faces can vary as a function of the trustworthiness of the web page on which it is shown.

This thesis suggests that web designers not rely on a list of more trustworthy colours, but, as in the latest Sony web site design, use white and grey with small flashes of colours, such as blue or red. As well, the addition of some faces to a poorly designed web page can augment its trustworthiness, but it will still not be as trustworthy as a well-designed page with no face.

References

- AAI. (2012). *Adult Attachment Interview Protocol*. Retrieved March 24, 2012 from www.psychology.sunysb.edu/attachment/measures/content/aa_i_interview.pdf
- Ainsworth, M. D. S. *Infancy in Uganda: infant care and the growth of love*. Baltimore: Johns Hopkins University Press, 1967.
- Ainsworth, M.D.S., & Bell, S. (1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child Development*, 41, 49–67.
- Ainsworth, M.D.S. (1979). Infant-mother attachment. *American Psychologist*, 34 (10), 932-937.
- Ainsworth, M.D.S. (1989). Attachments beyond infancy. *American Psychologist*, 44, 709–716.
- Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.

- APA (1954). American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201-238.
- APA (1966) American Psychological Association. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Author.
- Aronson, E. (1972). *The Social Animal*. New York. The Viking Press.
- Asendorpf, J.B. (1987). Videotape reconstruction of emotions and cognitions related to shyness. *Journal of Personality and Social Psychology*. 53, 542-549.
- Asendorpf, J.B. (1989). Shyness as a final common pathway for two different kinds of inhibition. *Journal of Personality and Social Psychology*, 57, 542-549.
- Badre, A.N. (2000). The effects of cross cultural interface design orientation on world wide web user performance. GVE Research Technical Reports. Retrieved March 24, 20-12 from <http://www.caslon.com.au/webdesignguide6.htm>
- Barber, W., & Badre, A.N. (1998). *Culturability: The Merging of Culture and Usability*. 4th Conference on Human Factors and the Web. Basking Ridge, NJ. Retrieved January 04, 2012 from <http://zing.ncsl.nist.gov/hfweb/att4/proceedings/barber/index.html>
- Bartholomew, K. (1990). Avoidance of intimacy. *Journal of Social and Personal Relationships*. 7, 147-178.

- Bartholomew, K., & Horowitz, L.M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology*. 61(2), 226-244.
- Battaglia, M., Ogliari, A., Zanoni, A., Villa, F., Citterio, A., Binaghi, F. & Fossati, A. (2004). Children's discrimination of expressions of emotions: Relationship with indices of social anxiety and shyness. *Journal of American Academy of Child Adolescent Psychiatry*, 43 (3), 358-365.
- Battaglia, M., Ogliari, A., Zanoni, A., Citterio, A., Pozzoli, U., Giorda, R., Maffei, C, & Marino, C. (2005). Influence of the serotonin transport promoter gene and shyness on children's cerebral responses to facial expressions. *Archives of General Psychiatry*, 62 (1), 85-94.
- Beidel, D.C., & Turner, S.M. (1998). *Shy Children, Phobic Adults: Nature & Treatment of Social Phobia*. Washington, DC: American Psychological Association.
- Berg, J.E., Dickhaut, J.W. & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*.10 (1), 122-42.
- Berger, T., Hohl, E., & Caspar, F. (2009). Internet-based treatment for social phobia: A randomized controlled trial. *Journal of Clinical Psychology* , 65(10), 1021-1035.
- Birren. F. (1982). *Light, colour, and environment*. New York. Van Nostrand Reinhold.
- Bonnardel, N., Piolat, A., Le Bigot,L.(2011). The impact of colour on Website appeal and users' cognitive processes, *Displays*, 32(2),69-80.

- Bortnick, K., Henderson, L., & Zimbardo, P. (2002). The Shy Q, a measure of chronic shyness: Associations with interpersonal motives, interpersonal values and self-conceptualizations. Poster presented at the 36th annual conference of the Association for the Advancement of Behavior Therapy (Nov14-17, 2002)
- Bowlby, J. (1946). *Forty-four juvenile thieves: Their characters and home-life*. London: Baillère, Tindall & Cox. First published in the *International Journal of Psycho-Analysis*, 25, 19–53, and 107-128.
- Bowlby J (1958). The nature of the child's tie to his mother. *International Journal of Psychoanalysis*, 39 (5), 350-373.
- Bowlby, J. (1982). *Attachment and loss, Vol. 1-3*. London: Hogarth.
- Bowlby, J. (1988). *A Secure Base*. New York: Basic Books.
- Bretherton, I. (1992). The origins of attachment theory: John Bowlby and Mary Ainsworth. *Developmental Psychology*, 28, 759-775.
- Buss, A.H. (1986). Two kinds of shyness. In R. Schwarzer (Ed.). *Self-related cognition in anxiety and motivation* (pp. 65-75). Hillsdale, NJ: Erlbaum.
- Carducci, B.J. (2002). *Shyness: A Bold New Approach*. New York: Harper Collins. (Original work published 1999)
- Carducci, B. J., & Zimbardo, P. G. (1995). Are you shy? *Psychology Today*. November/December, 28, 34-40, 64, 66, 68.
- CareFirst (2002). CareFirst Annual Report 2002. Retrieved Oct 06, 2011 from <https://www.carefirst.com/company/cf2002/SITE/color/index.html>

- Chavira, D., Stein, M., & Malcarne, V. (2002). Scrutinizing the relationship between shyness and social phobia. *Journal of Anxiety Disorders*, 16(6), 585-598.
- Cheek, J.M., & Buss, A.H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology*, 41, 330-339.
- Cheek, J.M. (1983). *13 item shyness scale*. Unpublished, Wellesley College, Wellesley MA 02181. Retrieved Oct 05 2011 from <http://www.wellesley.edu/Psychology/Cheek/research.html>
- Cheek, J.M., & Melichor, L.A. (1985). *Measuring the Three Components of Shyness*. In M.H. Davis & S.L. Franzoi (Co-chairs), *Emotion, Personality, and personal Well-Being II*. Symposium conducted at the annual convention of the American Psychological Association, Los Angeles. Retrieved November 01, 2011 from http://www.wellesley.edu/Psychology/Cheek/howshy20_text.html
- Cheskin Research and Studio Archetype/Sapient. (1999). *eCommerce Trust Study*. Retrieved Nov 1 01, 2011 from http://www.cheskin.com/cms/files/i/articles/17__report-eComm%20Trust1999.pdf
- Chesney, T., & Su, D.K.S. (2010). The impact of anonymity on weblog credibility. *International Journal of Human-Computer Studies*, 68, 710-718.
- Clark, M.C., & Payne, R.L. (1997). The nature and structure of workers' trust in management. *Journal of Organizational Behavior*, 18, 205-224.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 115, 155-159.

- Constantinople, A. (1969). An Eriksonian measure of personality development in college students. *Developmental Psychology*, 1(4), 357-372.
- Coopersmith, S. (1981). *The Coopersmith Self-esteem Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Cramer D. (1997). *Basic Statistics for Social Research: Step-By-Step Calculations and Computer Techniques Using Minitab*. New York. Routledge.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334.
- Cronbach, L.J., & Meehl, P.E. (1955) Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354.
- Crozier, W. R. (1995). Shyness and self-esteem in middle childhood. *British Journal of Educational Psychology*, 65(1), 85-95.
- Crozier, W. R. (2005). Measuring shyness: Analysis of the Revised Cheek and Buss Shyness Scale. *Personality & Individual Differences*, 38(8), 1947-1956.
- Cyr, D. (2008) Modeling web site design across cultures: relationships to trust, satisfaction, and e-loyalty. *Journal of Management Information Systems*. 24(4), 47-72.

- Cyr, D., Head, M., and Larios, H. (2010). Colour appeal in website design within and across cultures: a multi-method evaluation. *International Journal of Human-Computer studies*, 68: 1-21.
- Cyr, D., & Trevor-Smith, H. (2004). Localization of Web Design: An Empirical Comparison of German, Japanese, and U.S. Website Characteristics. *Journal of the American Society for Information Science and Technology*, 55(13), 1-10.
- Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, 2, 265-279.
- Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, 13, 123-139.
- DSM-IV. (1994). American Psychiatric Association. *Diagnostic and Statistical Manual Of Mental Disorder*. (4th ed.). Washington, DC: American Psychiatric Association.
- Duggan, E.S. & Brennan, K.A. (1994) Social avoidance and its relation to Bartholomew's adult attachment typology. *Journal of Social and Personal relationships*. 11, 147-153.
- Duncker, E., Theng, Y.L., & Mohd-Nasir, N. (2000). Cultural usability in digital libraries. *Bulletin of the American Society for Information Science* 26(4). Retrieved January 04, 2012 from http://www.asis.org/Bulletin/May-00/duncker_et_al.html

- Eagly, A.H., Ashmore, R.D., Makhijani, M.G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*. 110(1), 109-128.
- Ebeling-White, S., Frank, M.L., & Lester, D. (2007). Shyness, internet use, and personality. *CyberPsychology & Behavior*, 10(5), 713-716.
- Elliot, A.J.; Covington, M.V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, 13(2), 73-92.
- Elliot; A.J., Maier, M.A. (2007). Color and psychological functioning *Current Directions in Psychological Science*, 16 (5), pg. 250-254.
- Eysenck, H.J., & Eysenck, S.B.G. (1968). *Manual for the Eysenck Personality Scales*. San Diego, CA: Educational and Industrial Testing Service.
- Eysenck, H.J., & Eysenck, S.B.G. (1991). *Manual of the Eysenck Personality Scales*. London: Hodder & Stoughton.
- Feingold, A. (1992) Good-looking people are not what we think. *Psychological Bulletin*. 111(2), 304-341.
- 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. *Proceedings of ACM CHI 2001. Extended Abstracts*. Minneapolis, MN, US, April 20-25, 2001, pp 295-296.

- Fogg, B.J., & Tseng, H. (1999). The elements of computer credibility. *Proceedings of ACM CHI 99 Conference on Human Factors in Computing Systems. 1: 80-87* NY. ACM Press. Retrieved March 12, 2006 from <http://www.acm.org/pubs/articles/proceedings/chi/302979/p80-fogg/p-80fogg.pdf>
- Fogg, B.J., Soohoo, C., Danielson, D., Marable, L., Stanford, J., & Tauber, E.R. (2002). How do people evaluate a Web site's credibility? Retrieved March 24, 2012 from <http://credibility.stanford.edu/publications.html>
- Fogg, B.J., Soohoo, C., Danielson, D., Marable, L., Stanford, J., & Tauber, E.R. (2003). How do users evaluate the credibility of Web sites? A study with over 2500 participants. *Proceedings of the 2003 conference on Designing for User Experience: ACM. 1-15.*
- Fogg, B.J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. *Proceedings of CHI 2003.* ACM Press.
- Fogg, B.J., Kameda, T., Boyd, J., Marshall, J., Sethi, R., Sockol, M. & Trowbridge, T. (2002). Stanford-Makovsky Web Credibility Study 2002: Investigating what makes Web sites credible today. A Research Report by the Stanford Persuasive Technology Lab and Makovsky & Company. Stanford University. Retrieved Oct 08, 2011 from [www. Webcredibility.org](http://www.Webcredibility.org)
- Friedman, M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association. 32,675-701.*
- Friedman, H.H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management. 9(3), 114-123.*

- Friedman, H.H., Linda W. Friedman, L.W., & Gluck, B. (1988). The effects of scale-checking styles on responses to a semantic differential Scale, *Journal of the Market Research Society*, 30(4), 477-481.
- Galitz, W.O (1997). *The essential guide to user interface design: An introduction to GUI design principles and techniques*. New York. John Wiley & Sons.
- Gambetta, Diego (2000) 'Can We Trust Trust?', in Gambetta, Diego (ed.) *Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213-237, available at <http://www.sociology.ox.ac.uk/papers/gambetta213-237.doc>
- Gilkinson, H. (1942). Social fears as reported by students in college speech classes. *Speech Monographs*, 9, 141-160.
- Girodo, M. (1978). *Shy?* New York: Pocket Books.
- Gorn, G.J., Chattopadhyay, A., Yi, T., & Dahl, D.W. (1997). Effects of color as an executional cue in advertising: They're in the shade. *Management of Science*, 43 (10), 1387-1400.
- Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw-Hill.
- Guttman, L. (1954). An outline of some new methodology for social research. *Public Opinion Quarterly*. 18, 395-404.
- Guttman, R, & Greenbaum, C.W. (1998). Facet theory: Its development and current status. *European Psychologist*, 3(1), 13-36.
- Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1, 69-97.

- Hardin, M.G., Perez-Edgar, K., Guyer, A.E., Pine, D.S., Fox, N.A., & Ernst, M. (2006). Reward and punishment sensitivity in shy and least-shy adults: Relations between social and motivated behavior. *Personality and Individual Differences, 40*, 699-711.
- Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction, 19*, 319-349.
- Heiser, N.A., Turner, S.M., & Beidel, D.C. (2002). Shyness: Relationship to social phobia and other psychiatric disorders. *Behaviour Research and Therapy, 41*, 209-221.
- Henderson, L., & Zimbardo, P.G. (2001). Shyness, social anxiety, and social phobia. In S.G. Hofman & P. M. DiBartolo (Eds.), *From social anxiety to social phobia* (pp. 46-64). Needham Heights, MA: Allyn & Bacon.
- Hopko, D.R., Stowell, J., Jones, W.H., Armento, M.E.A., & Cheek, J.M. (2005). Psychometric properties of the Revised Cheek and Buss Shyness Scale. *Journal of Personality Assessment, 84*(2), 185-192.
- Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureno, G., & Vilasenor, V.S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*(6), 885-892.
- Horowitz, L.M., Rosenberg, S.E., & Bartholomew, K. (1993). Interpersonal problems, attachment styles, and outcome in brief dynamic psychotherapy. *Journal of Consulting and Clinical Psychology, 61*(4), 549-560.
- Ireland, S.R., Warren, Y.M., & Herring, L.G. (1992). Anxiety and colour saturation preference. *Perceptual and Motor Skills, 75*, 545-546.

- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1217-1218.
- Jamieson, S. (2005). Author's reply. *Medical Education*, 39, 971.
- Jarvis, B. G. (2008). MediaLab (Version 2008.1.33) [Computer Software]. New York, NY: Empirisoft Corporation.
- Jacobs, K.W. &Hustmyer, F.E. Jr. (1974). Effects of four psychological primary colors on gsr, heart rate and respiratory rate. *Perceptual and Motor Skills*, 38, 763-766.
- Jian, J-Y., Bisantz, A.M., & Drury, C.G. (1998). Towards an empirically determined scale of trust in computerized systems: Distinguishing concepts and types of trust. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. Chicago, IL. Human Factors and Ergonomics Society. Santa Monica, CA. 501-505.
- Jian, J-Y., Bisantz, A.M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
- Jones, W.H., & Briggs, S.R.(1986). *Manual for the Social Reticence Scale*. Palo Alto, CA: Consulting Psychologists Press.
- Jones, W.H., Briggs, S.R., & Smith, T.G. (1986). Shyness: Conceptualization and measurement. *Journal of Personality and Social Psychology*, 51(3), 629-639.
- Jones, G. R., & George, J.M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23(3), 531-546.

- Jung, C.G. (1923). *Psychological Types*. London: Routledge & Kegan Paul.
- Kagan, J., & Snidman, N. (1999). Early childhood predictors of adult anxiety disorders. *Biological Psychiatry*, 46 (11), 1536-1541.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*. 38(4), 319-342.
- Kendall, M. G., Babington Smith, B. (1939). The Problem of m rankings. *The Annals of Mathematical Statistics* 10 (3): 275–287
- Kim, J., Lee, J., & Choi, D. (2003). Designing emotionally evocative homepages: An empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies*. 59, 899-940.
- Kim, J., & Moon, J-Y. (1998) Designing towards emotional usability in customer interfaces-trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, 10, 1-29.
- Kobo (2011) Retrieved Oct 05, 2011 from <http://www.cbc.ca/news/technology/story/2010/04/08/f-kobo-ereader.html>
- Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*. 126(3), 390-423.
- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 6, 269-298.

- Leary, M.R. (1983a) .Social Anxiousness: The construct and its measurement. *Journal of Personality Assessment*, 47, 66-75.
- Leary, M. R. (1983b). The conceptual distinctions are important: another look at communication apprehension and other related constructs. *Human Communication Research*, 10, 65-75.
- Leary, M. R. (1986). Affective and behavioural components of shyness: Implications for theory, measurement, and research. In Jones, W.H., Cheek, J. M., & Briggs, S.R. (Eds.). *Shyness: Perspectives on research and treatment* (pp. 27-38). New York: Plenum.
- Leary, M.R. (1991). Social anxiety, shyness, and related constructs. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.). *Measures of Personality and Social Psychological Attitudes*: 161-194. New York: Academic Press.
- Leary, M.R., Kowalski, R.M., & Campbell, C.D. (1988). Self-presentational concerns and social anxiety: The role of generalized impression expectancies. *Journal of Research in Personality*, 22, 308-321.
- Lewicki, R.J., McAllister, D.J., & Bies, R.J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, 23(3), 438-458.
- Lewin, K. (1948). *Resolving social conflicts: selected papers on group dynamics*. Gertrude W. Lewin (Ed.). New York: Harper & Row.
- Li, J., & Chignell, M. (2010) Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies*, 68, 589-602.

- Lichtlé, M.C. (2007) The effect of an advertisement's color on emotions evoked by an ad and attitude toward the ad: the moderating role of the optimal stimulation level, *International Journal of Advertising*, 26(1), 32-62.
- Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention Web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25(2), 115-126.
- Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P. (2011). An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction (TOCHI)*. 18(1): 1-30.
- Madden, T.J., Hewett, K., & Roth, M.S. (2000). Managing images in different culture: Across-national study of colour and preferences. *Journal of International Marketing*, 8(4), 90-107.
- Main, Mary. (1996). Introduction to the special section on attachment and psychopathology: 2. Overview of the field of attachment. *Journal of Consulting and Clinical Psychology*, 64(2), 237-243.
- Main M (1999). Epilogue: Attachment Theory: Eighteen Points with Suggestions for Future Studies". In Cassidy J, Shaver PR. *Handbook of Attachment: Theory, Research and Clinical Applications*. New York: Guilford Press. 845–887

- Main, M., & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened/frightening parental behavior the linking mechanism? In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the Preschool Years: Theory, Research, and Intervention*, 161-182. Chicago, IL: University of Chicago Press.
- Main, M., & Solomon, J. (1986). Discovery of an insecure-disorganized/disoriented attachment pattern: Procedures, findings and implications for the classification of behavior. In T. B. Brazelton & M. W. Yogman (Eds) *Affective development in infancy* (pp. 95-124) Westport, CT, US: Ablex Pub.
- Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. In M. T. Greenberg, D. Cicchetti & E. M. Cummings (Eds.), *Attachment during the preschool years: Theory, research and intervention*. (pp. 121–160). Chicago: University of Chicago Press.
- Mancini, P. (1993). Between trust and suspicion: How political journalists solve the dilemma. *European Journal of Communications*, 8, 33-51.
- Mann, H.B., & Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.
- Masters, J.R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement*, 1(1), 49-53.
- McCroskey, J.C. (1970). Measures of communication-bound anxiety. *Speech Monographs*. 37(4), 269-277.

- McCroskey, J.C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communications Research*, 4, 78-96.
- McCroskey, J.C., & Anderson, J.F., Richmond, V.P., & Wheelless, L.R. (1981). Communication apprehension of elementary and secondary students and teachers. *Communication Education*. 30, 122-132.
- McCroskey, J.C., & Richmond, J.C. (1982). Communication apprehension and shyness: Conceptual and operational distinctions. *Central States Speech Journal*. 33(3), 458-468.
- McKnight, D.H., Cummings, L.L., & Chervany, N.L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, 23(3), 473-490.
- McKnight, D.H., & Chervany, N.L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.): *Trust in cyber-societies* (pp. 27-54). Berlin. Springer-Verlag.
- McKnight, D.H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334-359.
- McKnight, D.H., Kacmar, C.J., & Choudhury, V. (2004). Dispositional trust and distrust distinctions in predicting high- and low-risk internet expert advice site perceptions. *E-Service Journal*, 3(2): 35-55.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.

- Miller, R.S. (1995). On the nature of embarrassability: Shyness, social evaluation and social skill. *Journal of Personality*, 63, 315-339.
- Nichols, A. L., & Maner, J. K. (2008). The good subject effect: Investigating participant demand characteristics. *Journal of General Psychology*, 135, 151-165.
- Nielsen, J. (2006) Eyetracking research. Retrieved March 27, 2007 from <http://www.useit.com/eyetracking/>
- Nisbett, R.E. & Wilson, T.D. (1977a). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3): 231-259.
- Nisbett, R.E. & Wilson, T.D. (1977b). The halo effect: Evidence for unconscious alteration of judgements. *Journal of Personality and Social Psychology*, 35(4): 250-256.
- Noiwan, J. & Norcio, A. F. (2006) Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. *International Journal of Human-Computer Studies*, 64, (2), 103-122.
- Norman, Geoff. 2010. Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Sciences Education*, 15, 625-632.
- Ohanian, R. (1990) Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*. 19 (3): 39-52.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 17 (11): 776-783.

- Otto, M.W., & Safren, S.A. (2001). Mechanisms of action in the treatment of social phobia. In S. G. Hofmann & P.M. DiBartolo (Eds.). *From social anxiety to social phobia* (pp. 391-407). Needham Heights, MA: Allyn & Bacon.
- OED (2012). Oxford University Press retrieved March 23, 2012 from <http://www.oed.com/view/Entry/179089?rskey=u46lik&result=3&isAdvanced=false>.
- Pell, G. (2005). Use and misuse of Likert scales. *Medical Education*, 39, 970.
- Pew Internet and American Life Project. (2009). Retrieved Oct 05, 2011 from <http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information/10-Depression-anxiety-stress-or-mental-health-issues/1-Information-about-depression-anxiety-stress-or-mental-health-issues.aspx>
- Phan, K.L., Fitzgerald, D.A., Nathan, P.J., & Tancer, M.E. (2006). Association between amygdala hyperactivity to harsh faces and severity of social anxiety in generalized social phobia. *Biological Psychiatry*, 59, 424-429.
- Phillips, G.M. (1968). Reticence: Pathology of the normal speaker. *Speech Monographs*, 35, 39-49.
- Phillips, G.M., & Metzger, N.J. (1973). The reticent speaker: Etiology and treatment. *Journal of Communication Disorders*, 6(1), 11-28.
- Pozzulo, J.D., Coplan, R.J., & Wilson, J. (2005). Not too 'shy' to help catch a thief: Recall memory of shy versus least-shy. *Personality and Individual Differences*, 39(8), 1361-1372.

- Prakash, K., & Coplan, R.J. (2003). Shy skaters? Shyness, coping, and adjustment outcomes in female adolescent figure skaters. *Athletic Insight*, 5, 1-19.
- Princeton Survey Research Associates (2002). *A Matter of Trust: What Users Want From Web Sites*. Results of a National Survey of Internet Users for Consumer WebWatch. Retrieved Oct 06, 2011 from <http://www.consumerWebwatch.org/>
- Prior V, & Glaser D (2006). *Understanding Attachment and Attachment Disorders: Theory, Evidence and Practice*. Child and Adolescent Mental Health, RCPRTU. London and Philadelphia: Jessica Kingsley Publishers.
- Rapp, M. (1999). Communication apprehension, writing apprehension, introversion, and computer-mediated communication technology use in education: A call for research. *Society for Information Technology and Teacher Education International Conference*, 1280-12
- Reips, U.-D., & Funke, F. (2008) Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods*, 40(3), 699-704.
- Rempel, J.K., Holmes, J.G., & Zanna, M.P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*. 49(1), 95-112.
- Reynolds, W.M. (1982). Development of reliable and valid short forms of the Marlowe-Crown Social Desirability scale. *Journal of Clinical Psychology*, 38(1), 119-125.

- Riegelsberger, J., & Sasse, M.A. (2002). Face it-Photos don't make a web site trustworthy. *Proceedings of CHI 2002. Extended Abstracts*. Minneapolis, MN, US, April 20-25, 2002, pp. 742-743.
- Riegelsberger, J., Sasse, M.A., & McCarthy, J.D. (2003). Shiny happy people building trust? Photos on e-commerce Websites and consumer trust. *Proceedings of CHI 2003, 5(1), 121-128*.
- Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (Eds). (1991). *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press
- Rosenberg, Morris. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.
- Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35, 651-665*.
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26, 443-452*.
- Rotter, J. B. (1980). Interpersonal trust, trustworthiness and gullibility. *American Psychologist, 35(1), 1-7*.
- Rozin, P. (1986). One trial learning of acquired likes and dislikes in humans: Disgust as a US, food predominance and negative learning predominance. *Learning and Motivation, 17, 180-189*.
- Ryan, T.A. (1960). Multiple comparisons in psychological research. *Psychological Bulletin, 56(1), 26-47*.
- Schaffer, H. R. (1971). *The growth of sociability*. London: Penguin Book.

- Schank, R., & Neaman, A. (2001). Motivation & failure in educational simulation design. In K.D. Forbus and P.J. Feltovich (Eds.). *Smart Machines in Education*. Menlo Park, CA: MIT Press (pp 37-70)
- Schmidt, L., & Robinson, T.N. (1992). Low self-esteem in differentiating fearful and self-conscious forms of shyness. *Psychological-Reports*, 70 (1), 255-257
- Shepard, L.A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 405-450). Washington, DC: American Educational Research Association.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: John Wiley.
- Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004). Trust and mistrust of online health sites. CHI 2004.
- Sillence, E., Briggs, P., Harris, P. & Fishwick, L. (2006). A framework for understanding trust factors in web-based health advice. *International Journal of Human-Computer Studies*, 64, 697-713.
- Sillence, E., Briggs, P., Harris, P. & Fishwick, L. (2007). How do patients evaluate and make use of online health information? *Social Science and Medicine*. 64, 1853-1862..
- Simonian, S.J., Beidel, D.C., Turner, S.M., Berkes, J.L., & Long, J.H. (2001). Recognition of facial affect by children and adolescents diagnosed with social phobia. *Child Psychiatry and Human Development*, 32(2), 137-145.

Sites (2011). These sites were accessible as of October 05, 2011.

<http://www.stoptheadddiction.com/>, <http://www.midwestcenter.com> ;

<https://www.shynesscurve.com/curve/index.asp>, healsocialanxiety.com,
shyness.com

Steinbruck, U., Schaumburg, H., Duda, S, & Kruger, T. (2002). A picture says more than a thousand words-Photographs as trust builders in e-commerce websites. In *Proceedings of CHI 2002: Extended Abstracts*, 748-749.

Strickland, L.S. (1958). Surveillance and trust. *Journal of Personality*, 26, 200-215

Strizke, W.G.K., Nguyen, A. & Durkin, K. (2004). Shyness and computer-mediated communication: A self-presentational theory perspective. *Media Psychology*, 6(1), 1-22.

Turk, C.L., Lerner, J., Heimberg, R.G., & Rapee, R.M. (2001). An integrated cognitive-behavioural model of social anxiety. In S.G. Hofmann & P.M. DiBartolo (Eds.). *From social anxiety to social phobia (pp 281-303)*. Needham Heights, MA: Allyn & Bacon.

Turner, S.M., Beidel, D.C., & Townsley, R.M. (1990). Social phobia: Relationship to shyness. *Behaviour Research and Therapy*. 28, 497-505.

Utz, S. (2000). Social information processing in MUDs: The development of friendships in virtual worlds. *Journal of Online Behavior*, 1(1). Retrieved May 06, 2012, from <http://www.behavior.net/JOB/v1n1/utz.html>.

Valdez P, Mehrabian A. (1994) Effects of color on emotions. *Journal of Experimental Psychology: General*. 123(4), 394-409.

- van der Horst, F.C., LeRoy, H.A., & van der Veer, R. (2008). "When Strangers Meet": John Bowlby and Harry Harlow on Attachment Behavior. *Integrative Psychological & Behavioral Science*, 42:370–388.
- Veljaca, K., & Rapee, R.M. (1998). Detection of negative and positive audience behaviours by socially anxious subjects. *Behaviour Research and Therapy*, 36, 311-321.
- Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33, 448-457.
- Wells, A., Clark, D.M., Salkovskis, P., Ludgate, J., Hackmann, A., & Gelder, M.G. (1995). Social phobia: The role of in-situation safety behaviors in maintaining anxiety and negative beliefs. *Behavior Therapy*, 26, 153-161.
- Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956-972.
- Wiegersma, S., & Van der Elst. (1988). "Blue phenomenon": Spontaneity or preference. *Perceptual & Motor Skill*, 66, 308-310.
- Wilcoxon, F. (1945). Individual comparisons in ranking methods. *Biometrics Bulletin*, 1, 80-83.
- Wilson, G.D. (1966). Arousal properties of red versus green. *Perceptual and Motor Skills*, 23, 947-949.
- Wilson, R.E. and Eckel, C.C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*. 59 (2):189-202.

- Winton, E.C., Clark, D.M., & Edelmann, R.J. (1995). Social anxiety, fear of negative evaluation and the detection of negative emotion in others. *Behaviour Research and Therapy* 33(2), 193-196.
- Wrightsmann, L.S. (1991). Interpersonal trust and attitudes toward human nature. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsmann (Eds.), *Measures of personality and social psychological attitudes* (pp. 373-412). San Diego, CA: Academic Press
- Yerkes, R.M., & Dodson, J.D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Zahorik, D. (1979). Learned changes in preferences for chemical stimuli: Asymmetrical effects of positive and negative consequences, and species differences in learning. In J. H. A. Kroeze (Ed.), *Preference behavior and chemoreception*. London: Information Retrieval.
- Zajonc, R.B. (1965). Social facilitation. *Science*, 149, 269-274.
- Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*. 35 (2), 151-175.
- Zimbardo, P.G., Ebbessen, E.B., & Maslach, C. (1977). *Influencing Attitudes and Changing Behaviour*. 2nd ed. Addison-Wesley: Reading, Mass.
- Zimbardo, P.G. (1989). *Shyness*. Berkley Publishing Group. New York. (Original work published 1977).
- Zimbardo, P.G., & Radl, S. (1999). *The Shy Child*. Cambridge, MA: McGraw-Hill. (Original work published 1981).

Zimmerman, D.W. & Zumbo, B.D. (1993). Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *Journal of Experimental Education*, 62 (1), 75-86.

Zook, A. and Sippes, G. 1985. Cross-validation of a short form of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 41(2):236-238.

Appendix A

	Results of a Google search, September 22, 2007, under <healthy living>	Study 1 pages
1	http://www.weightwatchers.ca/health/hwb/index.aspx	Page 4
2	http://fatloss4idiots.com/	
3	http://www.articleworld.biz/topics/23/healthy%20living.aspx	Page 6
4	http://www.nlm.nih.gov/medlineplus/healthyliving.html	
5	http://www.apta.org/AM/Template.cfm?Section=Healthy_Living&Template=/TaggedSite/TaggedSiteDisplay.cfm&TPLID=213&ContentID=20858	
6	http://meriter.staywellsolutionsonline.com/	
7	http://www.kraftfoods.com/kf/HealthyLiving/	
8	http://www.healthyliving.org/	Page 2
9	http://www.nutritiondata.com/	
10	http://healthylivinginfo.com/	
11	http://www.healthylivingmarket.com/	Page 1
12	http://www.gov.mb.ca/healthyliving/	Page 3
13	http://www.healthyliving.gov.uk/	
14	http://diabeticgourmet.com/Healthy_Living/	Page 8
15	http://www.mayoclinic.com/health/HealthyLivingIndex/HealthyLivingIndex	
16	http://www.lifescrypt.com/channels/healthy_living/index.asp	
17	http://www.medicinenet.com/healthy_living/article.htm	
18	http://www.cdc.gov/HealthyLiving	
19	http://healthy.net/	Page 5
20	http://www.healthy-living.org/	Page 7

Appendix B

16-item scale based on factors from Sillence et al (2004, 2006, 2007) and Fogg, Soohoo et al (2002).

1. This Web site name is

Inappropriate 1.....2.....3.....4.....5.....6.....7 Appropriate

2. This Web site layout is

Simple 1.....2.....3.....4.....5.....6.....7 Complex

3. This Web site layout is

Not busy 1.....2.....3.....4.....5.....6.....7 Busy

4. This Web site layout is

Not clear 1.....2.....3.....4.....5.....6.....7 Clear

5. This Web site navigation is

Not obvious 1.....2.....3.....4.....5.....6.....7 Obvious

6. This Web site has too many menus

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

7. This Web site design is

Boring 1.....2.....3.....4.....5.....6.....7 Not boring

8. This Web site's use of colour is

Good 1.....2.....3.....4.....5.....6.....7 Not good

9. This Web site has too much advertising.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

10. The print size on this Web site is too small.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

Appendix B (continued)

11. There is too much text on this Web site.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

12. This Web site has a corporate look and feel.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

13. This Web site looks professional.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

14. This Web site looks organized.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

15. This Web site makes good use of graphics.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

16. This Web site has a familiar feel to it.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

Appendix C

Friedman, Kendall, & Wilcoxon, for Pilot Study n=10

Pilot Study	n=10
Friedman	test
total site	mean rank
K11	18.45
H08	16.40
L12	14.75
E05	14.00
P16	13.20
A01	13.15
B02	13.10
J10	12.90
O15	10.75
M13	10.35
D04	10.25
S19	9.75
G07	9.55
F06	9.45
I09	8.80
R18	7.05
Q17	6.05
C03	5.85
T20	3.15
N14	3.05

Friedman	Chi-Square	92.810
	df	19
	Asymp. Sig.	1.05e-11

Kendall's Coefficient of Concordance	
W= 0.4884755	

Wilcoxon	L12 vs C03
	negative ranks 8
	positive ranks 2
	ties 0
	Z= -2.095
sig. (2-tailed)	0.036

Appendix D

Spearman Correlations of Rank Orders by Question for 8 Web pages

Pilot Study for Study 1 n=10

item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	name	simple	busy	clear	obvious	menus	boring	colour	ads	print size	amount of text	corporate	professional	organized	graphics	familiar
1	1.00						.79*						.74*	.81*	0.81*	.90*
2		1.00	.88**			.86**					.83*					
3			1.00	.71*		.95**					.98**					
4				1.00	.83*	.74*		.83*	.83*		.81*			.76*		.73*
5					1.00	.83*			.74*		.79*					
6						1.00					.98**					
7							1.00	.93**					.79*	.92**	.99**	.84**
8								1.00	.74*					.91**	.91**	.79*
9									1.00	.79*						
10										1.00						
11											1.00					
12												1.00	.91**			
13													1.00	.88**	.83*	.87**
14														1.00	.93**	.96**
15															1.00	.85**
16																1.00

* significant at p<.05

**significant at p<.01

Appendix E

Pilot Study for Study 1

The purpose of this pilot study was to select eight Web pages from an original group of 20 Web pages (Appendix A). Using eight pages rather than 20 in Study #1 would increase its statistical power. Ryan (1960) portions the “alpha” levels for multiple comparisons (probability of a Type 1 error) using the formula $\frac{2\alpha}{k(r-1)}$, where k = the total number of treatment levels, and r = the number of levels between the treatments being tested. Using smaller values for k and r allows for a greater chance of finding a significant result.

As well, this pilot study aimed to pare down the number of items in the scale to a smaller number.

*Method**Participants*

Ten participants, 4 male and 6 female, took part in this pilot study. All were in their late teens or early twenties, 1st and 2nd year undergraduates at Carleton University, except two non-students in their mid 40's. The students were recruited through an electronic bulletin (Appendix F) posted on SONA, the Carleton University Psychology Experiment Sign-Up System. The students were recompensed with 1% added to their Psychology course mark. The non-students were not recompensed. No evaluation of shyness was necessary in this study. Each session lasted approximately 45 minutes.

Appendix E (continued)

Apparatus

The laptop computer used was a Toshiba Satellite A200 AH9, 1.73 GHz Intel CPU, with 2 GB RAM. The monitor was set at 32 bits, 1280 x 800 pixel screen resolution, 60 Hz refresh rate, with screen settings: Gamma 1.0, Brightness 0, Contrast 50. JPG images, at 720 x 563 pixel resolution in 32-bit colour, were presented as screen shots for each Web page as they would appear in an Internet Explorer browser window. The Web pages were viewed in a windowless room, 6m by 3m by 2.5 m high, illuminated by three sets of two fluorescent bulbs (4-foot Sylvania Octron 4100k, 32 Watts each). The laptop screen was leaning away from the participants at an angle of approximately 30° to the vertical. The screen was approximately 61 cm from the participant.

An application created in MediaLab v 2008.1.33 (Jarvis, 2008) was used to present Web pages and collect responses to the scales.

Materials

The materials included informed consent forms, debriefing forms, and pens. Home pages from the first 20 non-redundant Web sites were chosen from a Google search on “healthy living” (Appendix A). In order to approximate real-life conditions, I mimicked the behaviour of a Web surfer and chose home pages from the first 20 unique sites found from a Google search on “healthy living”. No other selection criteria were used, as the goal was to model the results that a user would generate in such an introductory search. The twenty Web pages were used as the treatment levels of the independent variable. The rating instrument (Appendix B)

Appendix E (continued)

was composed of 16 items based upon factors identified by Sillence et al. (2004, 2006, 2007) and Fogg, Soohoo et al., (2002) to describe Web sites that were perceived as trustworthy or credible by their respective participants.

Both Fogg, Soohoo et al. (2002) and Sillence et al. (2004, 2006, 2007) found the prominent aspects of a Web site at the initial stage of trust formation to be design, not content. Sillence et al (2006), asked their 40 participants, to identify the factors that led to the rejection or mistrust of Web sites. Of the 12 factors identified, six can be found in Fogg et al (2002). The 12 factors are: appropriate Web site name, complex layout, busy layout, clear layout, navigation, number of menus, boring design, use of colour, advertising, small print, too much text, corporate look and feel. These 12 factors, as well as four words from Fogg et al (2002), referring to professional, organized, graphics, and familiar brands, were chosen for the 16-item 7-point Semantic differential scale to be used in the pilot study.

Design

This pilot study employed a repeated measures design, with twenty treatment conditions, the twenty home pages from the Google search. The Web pages were presented randomly. There was one within subjects variable, the 16-item scale based upon factors identified by Sillence et al (2004, 2006) and Fogg, Soohoo et al (2002). (Appendix B). The responses to the within-subjects variable, the 16-item scale, were balanced, left to right, but not presented in random order. The coding for the responses to eight of the items were set so that “1” was least trustworthy, and

Appendix E (continued)

to the other eight so that “7” was scored as least trustworthy. If there was a fatigue effect so that the 16th item was evaluated differently from the first, this effect would be equal over the pages, and therefore, since the scores used were the total score by participant, not an issue for this study.

Procedure

To begin with, the participants were welcomed and asked to sign a consent form (Appendix G). Then, each participant was shown twenty Web site home page images, in a randomized order, on a laptop computer, as they would appear in an Internet Explorer browser window. The instructions were standardized by using an opening screen to explain the procedure (Appendix H). The screen explained that the participants could take their time looking at the images, and that a smaller replica of the image would be presented below each of the statements to be rated. After seeing each page for as long as they wished, the participant clicked a "Continue" button, and, while still viewing a smaller image of the page, responded to the 16 items, one at a time, using a 7-point semantic differential scale, displayed on the computer screen. The participants were told they could move the screen and/or the computer to a more comfortable viewing position if desired. Finally, the participants were debriefed and thanked (Appendix I).

Results

Because the number of participants was low, the total scores for each Web page were subjected to the non-parametric Friedman test (Friedman, 1937) to

Appendix E (continued)

generate an overall ranking (Appendix C). This generated a $\chi^2(19) = 92.810$, ($p = .000$) showing that the Web pages exhibited significant differences using the 16-item scale.

The Wilcoxon signed rank test (Wilcoxon, 1945) was then used to establish if significant differences existed between the third highest ranked Web page and the page ranked # 18. This would indicate that the set of top three and the set of bottom three were significantly different on this rating scale. This non-parametric repeated measures test resulted in a significance level of .036. Given the small size of the sample, this was considered a significant value. Kendall's W , which measures the level of agreement amongst the raters (Siegel and Castellan 1988), was also calculated and yielded a figure of .488, showing a modest, but highly significant ($p = 0.000$) amount of agreement among the participants.

Based on these findings, I chose the three highest-ranking pages, three lowest ranking, and two randomly chosen pages in the middle, for further analysis. This was done to increase the power of the next studies by diminishing the number of experimental stimuli (treatment levels) to be used. The eight Web pages were A01, C03, H08, K11, L12, N14, S19, and T20 (Appendix A).

Using the non-parametric Friedman test (Friedman, 1937), the eight Web pages, were ranked according to each item, to generate a set of 16 rankings. All of these rankings were significant although the orders differed in some respects.

The rankings of these eight pages, by each of the 16 items in the scale, were correlated using the Spearman non-parametric correlation function. It was thought

Appendix E (continued)

that, if certain items generated similar rankings, they likely tapped into the same concept, "trust". Certain rankings were highly correlated.

Item 1 "appropriate name" was highly correlated with item 7 "boring design" ($\rho = .790$), item 13 "professional look" ($\rho = .738$), item 14 "organized" ($\rho = .810$), item 15 "good graphics" ($\rho = .810$), and item 16 "familiar feel" ($\rho = .898$).

Item 4 "clear layout" correlated highly with item 8 ($\rho = .833$), item 14 ($\rho = .762$), and item 16 ($\rho = .731$).

Item 5 "obvious navigation" was retained because it is referenced specifically by Fogg (2002) under "Information Design/Structure" (28.1%).

Item 8 "colour" correlated highly with items 14 ($\rho = .905$), 15 ($\rho = .905$), and 16 ($\rho = .790$).

Item 10 "print size" was retained because it is mentioned specifically by Fogg (2002) stating that 46.1% of participants in the study assessed credibility of sites based on visual design specifically noting "typography" and "font size".

Finally, Item 12 "corporate look and feel" was retained because of a possible disparity between the findings of Sillence et al (2006, 2007) and this pilot study. Sillence et al (2006, 2007) indicated that their participants viewed "corporate look and feel" as having a negative relation to trust. Sillence et al (2006) report a sample of 30 women, and 10 men, but no other demographics are referenced. In Sillence et al (2004, 2007) the sample is 15 women. This seems to be one of the "four separate studies" making up Sillence et al (2006). So, 15 of the 40 had a mean age of 49

Appendix E (continued)

years (Sillence et al, 2004, 2007). The sample for this pilot study, on the other hand, was primarily university students in their late teens or early twenties. They may have had a different appreciation of corporations. As well Item 12 correlated highly with Item 13, “professional look” ($\rho = .905$). This item was determined to be of interest in the larger sample of Study 1.

As a result of this analysis, I chose to use the following 11 items for the next part of the thesis: Item 1 “appropriate name”; Item 4 “clear layout”; Item 5 “obvious navigation”; Item 7 “boring design”; Item 8 “good colour”; Item 10 “print size”; Item 12 “corporate look and feel”; Item 13 “professional look”; Item 14 “organized”; Item 15 “good graphics”; and Item 16 “familiar feel” (Appendix K).

These 11 items, using a 7-point scale, could be expected to produce a Cronbach reliability value (Cronbach, 1951) of about 78% (Masters, 1974).

Discussion

The purpose of this pilot study was to establish eight Web pages, three Web pages that rated as trustworthy, three that rated as least trustworthy, and two rated as having middling trustworthiness. This was accomplished.

As well, the sixteen original items had been narrowed down to eleven that seemed to target the concept of “trust” more carefully.

The pilot study data analysis suggested that certain items seemed to focus on the same concept, as they ranked the sites similarly. These items were, in fact, consistent with the factors identified by Sillence et al (2004, 2006, 2007).

Appendix E (continued)

Inappropriate name, layout, navigation, web design, use of colour, small print, and corporate look and feel were all retained in the 11-item scale. The 16-item scale used three layout items, *simple/complex, not busy/busy, and not clear/clear*, and the latter two did not seem to discriminate along the same lines as the 11 retained items. This could be simply because the eight sites that were kept were perceived to be equally well-designed with respect to these features, but not with respect to complexity. Similarly, *number of menus*, which was an interpretation by myself with respect to the category, *busy layout*, was not seen as a differentiator by the participants. Again, the items *too much advertising* and *too much text* could be good items for initial trust of a Web page generally, but these participants perceived these 8 sites to be equally well-designed with respect to these categories. The fact that the original 16 items, based on the findings of Sillence et al (2004, 2006, 2007) focused so well in the pilot study supports their research.

Although *corporate look and feel* did tap into the same concept as the other 15 items, the direction of its effect was surprising. The participants in Sillence et al (2004, 2006, 2007) regarded *corporate look and feel* as an indicator of mistrust. Supporting this view, comments in Fogg et al. (2002), under the category "Company Motive", were to the effect that the site "Seems too 'commercial'...", and "This site says to me ' Give us your money and get out.' ". Therefore, it was expected that the *corporate look and feel* item would be an indicator of distrust and initial rejection of a page. This was not the case for this 10-participant study.

Appendix E (continued)

Corporate look and feel, at least for this sample of young university students, was positively correlated with all of the other trust items. This may not be the case in another sample, or with respect to other Web pages. The 20 pages used in the pilot study may be non-threatening with respect to this factor for this sample. This word may have different connotations in different countries and for different demographics within a country. The groups tested in Sillence et al (2004, 2006, 2007) and Fogg et al. (2002) apparently assigned different meanings to this phrase than the participants in the pilot study.

Appendix F

Sample Recruitment Notice

(To be posted on the electronic recruitment bulletin)

**Web Credibility Study #1
(Photo Study #4)**

Hi, my name is Irvine Johnston and I am a Ph.D. student in the Human Oriented Technology Lab at Carleton University. I am conducting research to evaluate various Web sites (photos) based on various criteria (as to their perceived friendliness).

I am looking for people with any level of computer experience from novice to expert.

Your name will not be associated with the data and you will remain anonymous. The study will take about 45 minutes to complete and you will receive 1.0% towards your PSYC mark (plus my thanks!). The study will take place in HCI 3112C (go through the hallway on the south-west wall of Level 2 of Loeb; go through the doors at the end of the hall; turn left and go up the stairs to the 3d floor of the Human Computer Interaction building). For any additional information you can call me at _____, or e-mail me at ijohnst2@carleton.ca This study has been approved by the Carleton University Ethics Committee for Psychological Research.

Appendix G

Sample Informed Consent Form

Study 3

The purpose of an informed consent is to ensure that you understand the purpose of the study, and your involvement in the study. The informed consent must provide enough information so that you can determine whether or not you wish to participate in this study.

The purpose of this study is to study your perceptions of some Web sites (photos).

Research Personnel

The following people are involved in this research project and may be contacted at any time. Irvine Johnston: ijohnst2@connect.carleton.ca (Principal Investigator), 520-2600 ext. 6628, Dr. Gitte Lindgaard (Faculty Sponsor), 520-2600 ext. 2255, gitte_lindgaard@carleton.ca. If any ethical concerns about this study should arise please contact Dr. Monique Sénéchal (Chair, Carleton University Ethics Committee for Psychological Research), 613-520-2600 ext. 1155, monique_senechal@carleton.ca. Should you have any other concerns about this study, please contact Dr. Janet Mantler (Chair, Dept of Psychology), 613- 520-2600, ext. 4173, psychchair@carleton.ca

The procedure for this study is: To begin with, you will be asked to sign this consent form. Then, you will be asked to answer a series of questions concerning your perceptions of some Web sites. Next, you will be asked to fill out a short questionnaire. Finally, you will be debriefed and you will receive 1% towards your PSYC 1001, 1002, 2001 or 2002 mark.

The duration of this study will be about 45 minutes.

There are no potential risks or discomforts in this study and you will not be evaluated for right or wrong answers. You can take rest breaks throughout the study if you wish.

All participants' data will remain anonymous. The data collected will be coded such that your name will not be associated with the data. The data is to be used only by the researchers involved in this project.

You have the right to withdraw from the study at any time without penalty. You have the right to omit answering specific questions without penalty.

**This study has been approved by the Carleton University
Ethics Committee for Psychological Research.**

Statement of Agreement

I have read the above description of the study and understand the conditions of my participation. I agree to participate in this research project.

Date: _____

Participant Name: _____

Participant Signature: _____

Researcher Name: Irvine Johnston

Researcher Signature: _____

Appendix H

Sample Opening Instruction Screen for participants

WELCOME to this study of web pages

You are going to be presented with a series of 20 web pages, one at a time, and asked to rate each on various factors.

Each web page will be presented for as long as you like. When you are ready, left-click "CONTINUE" at the bottom right hand of the screen. The questions and the rating scales will appear, along with a smaller version of the web page.

You can rate each factor from 1 to 7, by using the number keys on the keyboard or left-clicking, using the mouse.

The wording of each scale is slightly different, so take a moment to read each statement carefully.

Once again, there are no right or wrong answers.

If you have any questions at this time, please ask the researcher for information.

Appendix I

Sample Debriefing Form

Thank you for participating in this experiment.

You were a participant in an experiment conducted by Irvine Johnston (Ph.D. candidate) and Dr. Gitte Lindgaard. There were no right or wrong answers. It is part of a multi-part study of a relatively unknown area of Web site design, namely: how the perceived trustworthiness and visual appeal of Web sites are affected by the use of various colours and photos

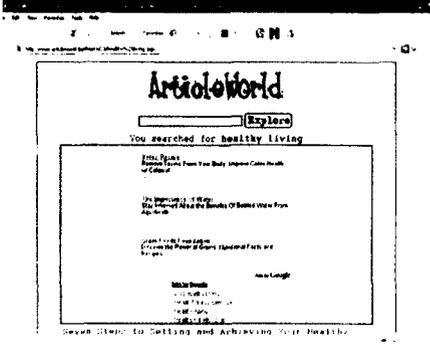
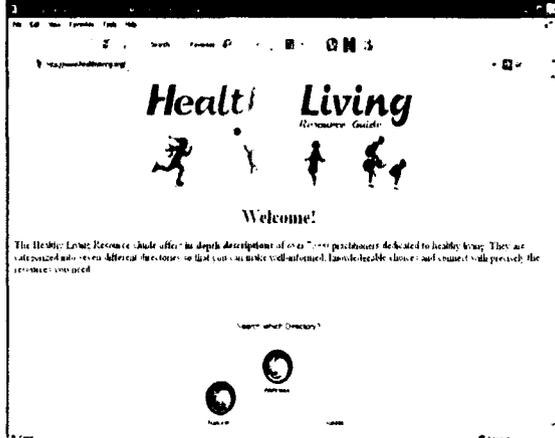
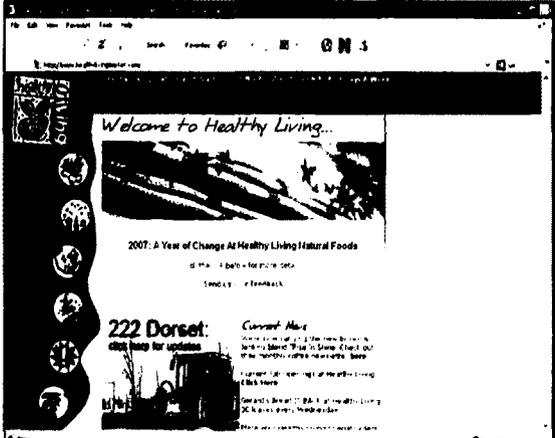
This study will provide experimental support for Web design principles. We hope that by learning more about Web design we can create better and more satisfactory software applications.

Thank you again for participating in the study. Your time and effort are greatly appreciated!

Please feel free to contact us at any time if you wish to discuss any aspects of the research. Irvine Johnston: ijohnst2@connect.carleton.ca (Principal Investigator), 520-2600 ext. 6628, Dr. Gitte Lindgaard (Faculty Sponsor), 520-2600 ext. 2255, gitte_lindgaard@carleton.ca. If any ethical concerns about this study should arise please contact Dr. Monique Sénéchal (Chair, Carleton University Ethics Committee for Psychological Research), 520-2600 ext. 1155, monique_senechal@carleton.ca. Should you have any other concerns about this study, please contact Dr. Janet Mantler (Chair, Dept of Psychology), 520-2600, ext. 4173, psychchair@carleton.ca.

Appendix J

8 Web Pages for Study1

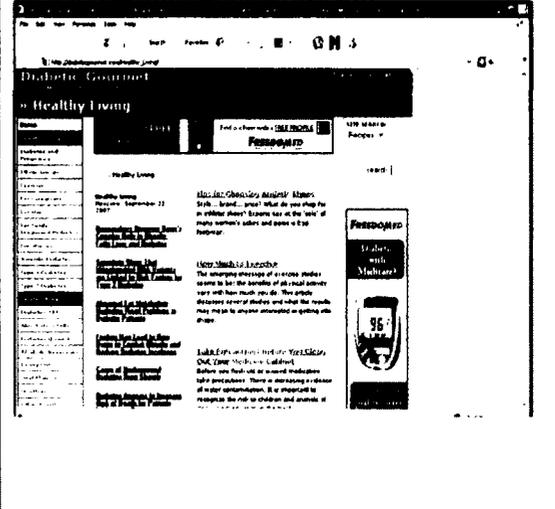
<p>A01. http://www.weightwatchers.ca/health/hwb/index.aspx</p> 	<p>C03. http://articleworld.biz/topics/23/healthy_living.aspx</p> 
<p>H08. http://www.healthyliving.org/</p> 	<p>K11. http://www.healthylivingmarket.com/</p> 

Appendix J(continued)

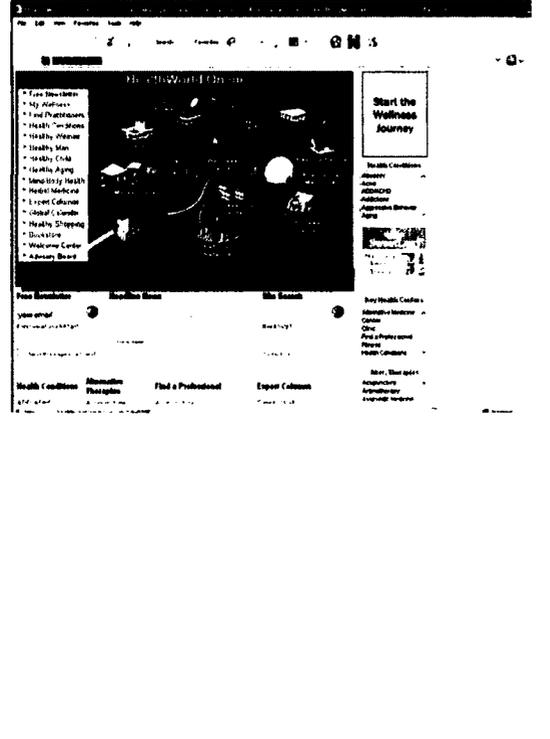
L12. <http://www.gov.mb.ca/healthyliving/>



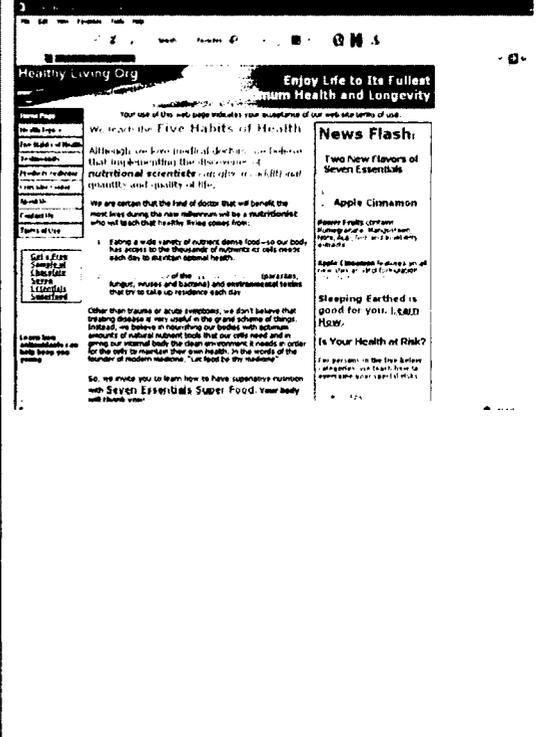
N14. http://diabeticgourmet.com/Healthy_Living/



S19. <http://healthy.net/>



T20. <http://www.healthy-living.org/>



Appendix K

The Adapted 11-Item Scale for Study 1

1. This Web site name is

Inappropriate	1.....2.....3.....4.....5.....6.....7	Appropriate
---------------	---------------------------------------	-------------

2. This Web site layout is

Not clear	1.....2.....3.....4.....5.....6.....7	Clear
-----------	---------------------------------------	-------

3. This Web site navigation is

Not obvious	1.....2.....3.....4.....5.....6.....7	Obvious
-------------	---------------------------------------	---------

4. This Web site design is

Boring	1.....2.....3.....4.....5.....6.....7	Not boring
--------	---------------------------------------	------------

5. This Web site's use of colour is

Good	1.....2.....3.....4.....5.....6.....7	Not good
------	---------------------------------------	----------

6. The print size on this Web site is too small.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

7. This Web site has a corporate look and feel.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

8. This Web site looks professional.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

9. This Web site looks organized.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

10. This Web site makes good use of graphics.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

11. This Web site has a familiar feel to it.

Disagree	1.....2.....3.....4.....5.....6.....7	Agree
----------	---------------------------------------	-------

Appendix L

Friedman, Kendall, and Wilcoxon Results for Study 1

	Mean Rank
A01 total	5.44
C03 total	2.29
H08 total	4.77
K11 total	5.85
L12 total	7.08
N14 total	3.82
S19 total	4.08
T20 total	2.66

	Mean Rank
L12 total	7.08
K11 total	5.85
A01 total	5.44
H08 total	4.77
S19 total	4.08
N14 total	3.82
T20 total	2.66
C03 total	2.29

Test Statistics

Friedman statistic	N	31
	Chi-Square	95.733
	df	7
	Asymp. Sig.	.000
Kendall's W	0.441	

Study 1 n=31, 8 sites, 11 Q's Wilcoxon

	L12	K11	A01	H08	S19	N14	T20	C03
L12		0.030	0.004*	0.000*	0.000*	0.000*	0.000*	0.000*
K11			0.490	0.012	0.004*	0.002*	0.000*	0.000*
A01				0.085	0.006*	0.002*	0.000*	0.000*
H08					0.178	0.090	0.001*	0.000*
S19						0.333	0.001*	0.000*
N14							0.028	0.015
T20								0.204
C03								

* significant at $\alpha = 0.05$ experiment-wises error rate (Ryan, 1960)

Appendix M

The Adaptations of the Jian and McKnight Trust Scales

The Jian Trust in a Machine Scale was originally designed to assess trust in a system. For the present purpose, the word "system" was replaced by "Web site" in nine of the 12 items in the scale. One of the statements "I am familiar with the system." implied some prior use of the site. Since I did not assume any prior experience with the eight Web sites, I omitted this statement completely. The other three statements were modified only slightly. The changes are tabled below.

Item	Jian items	Adapted scale
1	The system is deceptive.	This Web site is deceptive.
2	The system behaves in an underhanded manner..	There is something underhanded about this Web site
3	I am suspicious of the system's intent, action, or outputs.	I am suspicious of this Web site's intent.
4	I am wary of the system.	I am wary of This Web site.
5	The system's actions will have a harmful or injurious outcome.	If I used this Web site, its actions would have a harmful or injurious outcome.
6	I am confident in the system.	I am confident in This Web site.
7	The system provides security.	This Web site provides security.
8	The system has integrity.	This Web site has integrity.
9	The system is dependable.	This Web site is dependable.
10	This Web site is reliable.	This Web site is reliable.
11	I can trust the system.	I can trust This Web site.
	I am familiar with the system.	

Statement #2 was changed from "The system behaves in an underhanded manner" to "There is something underhanded about this web site", since a Web page, on initial viewing, cannot be said to "behave". Statement #3, "I am suspicious of the machine's intent, action, or output" became "I am suspicious of the intent of this web site". The words "action, or output" were deemed inappropriate for an

Appendix M (continued)

initial view and were omitted. Statement #5, “The system’s action will have a harmful or injurious outcome” also referred to the system’s “action”, which was not possible in an initial viewing. Therefore, a hypothetical statement was substituted: “If I used this Web site, its actions would have a harmful or injurious outcome”.

The statements in the adapted Jian scale, as in the original, were worded so that, for five, "trust" is to the left end of the scale, for six, "trust" is to the right end of the scale to control for any tendency to respond in a particular direction.

Eleven statements from the *Trusting Beliefs* and *Trusting Intentions* sections of the 68-item McKnight scale were modified and used to validate the Jian Trust scale. Statements irrelevant to the present study were omitted or modified. The changes are tabled below.

Item	McKnight items	Adapted scale
1	I believe that LegalAdvice.com would act in my best interests.	I believe that this Web site would act in my best interest.
2	If I required help, LegalAdvice.com would do its best to help me.	If I required help, this Web site would do its best to help me.
3	LegalAdvice.com is interested in my well-being, not just its own.	This Web site is interested in my well-being, not just its own.
4	LegalAdvice.com is truthful in its dealings with me.	This Web site is truthful in its dealings with me.
5	I would characterize LegalAdvice.com as honest.	I would characterize this Web site as honest.
6	LegalAdvice.com would keep its commitments.	This Web site would keep its commitments.
7	LegalAdvice.com is sincere and genuine.	This Web site is sincere and genuine.
8	When an important legal issue or problem arises, I would feel comfortable depending on the information provided by LegalAdvice.com.	I would feel comfortable using this Web site.
9	I can always rely on LegalAdvice.com in a tough legal situation.	I can always rely on this Web site.

Appendix M (continued)

Item	McKnight items	Adapted scale
10	I feel that I could count on LegalAdvice.com to help with a crucial legal problem.	I feel that I could count on this Web site.
11	I would not hesitate to use the landlord/tenant information LegalAdvice.com supplied me.	I would not hesitate to use this Web site.

Four of the 11 *Trusting Beliefs* items loaded on “competence”, and were omitted, since “competence” appears to be a synonym for what Fogg and Tseng (1999) called “expertise”, a second factor in credibility not an issue in this thesis. Since the McKnight scale focussed on an intention to use a specific Web site, LegalAdvice.com., the seven adapted statements simply replaced the phrase “LegalAdvice.com” with “this Web site”.

Of the 16 *Trusting Intentions* items, 12 were omitted: six referring to the expertise of the site, a second factor in credibility not an issue in this thesis; four referring to the possibility of paying to use the Web site, a choice which would apply to later stages in trust development (Sillence et al, 2006, 2007); and two that involved the words “confident”, and “secure”. The four adapted items therefore refer only to an affective relationship with the Web site and omit any reference to expertise. As an example, the statement “When an important legal issue or problem arises, I would feel comfortable depending on the *information* provided by LegalAdvice.com” was modified to “I would feel comfortable using this Web site”. The 11 adapted items from the *Trusting Intentions* and *Trusting Beliefs* sections used in the present research were judged to be indicators of the *initial trust*

Appendix M (continued)

in a Web site construct being studied. The McKnight statements were all structured so that "trust" is to the right end of the scale, the "agree" end.

Appendix N

The Abridged Marlowe-Crowne Social Desirability Scale

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true (T) or false (F) as it pertains to you personally. Circle the response (T or F) that applies.

1	It is sometimes hard for me to go on with my work if I am not encouraged.	1	T	(F)
2	I sometimes feel resentful when I don't get my way.	2	T	(F)
3	On a few occasions, I have given up doing something because I thought too little of my ability.	3	T	(F)
4	There have been times when I felt like rebelling against people in authority even though I knew they were right.	4	T	(F)
5	No matter who I'm talking to, I'm always a good listener.	5	(T)	F
6	There have been occasions when I took advantage of someone.	6	T	(F)
7	I'm always willing to admit when I make a mistake.	7	(T)	F
8	I sometimes try to get even rather than forgive and forget.	8	T	(F)
9	I am always courteous, even to people who are disagreeable.	9	(T)	F
10	I have never been irked when people expressed ideas very different from my own.	10	(T)	F
11	There have been times when I was quite jealous of the good fortune of others.	11	T	(F)
12	I am sometimes irritated by people who ask favors of me.	12	T	(F)
13	I have never deliberately said something that hurt someone's feelings.	13	(T)	F

Scoring Algorithm For each answer that matches the bracketed response above assign a value of 1. The total score can range from 13 (when all responses "match") to 0 (when no responses "match").

Reynolds, 1982

Appendix O

Pearson Product Moment Correlation of Jian Scale Items with Item #11

“I can Trust this Web Site”

Jian Scale Distrust-Oriented Items

	deceptive	underhanded	suspicious	wary	harmful
Page 4	-.70***	-.46**	-.78***	-.57***	-.35*
Page 6	-.42**	-.41**	-.41**	-.47**	-.26
Page 2	-.69***	-.65***	-.63***	-.73***	-.36*
Page 1	-.69***	-.58***	-.57***	-.60***	-.43**
Page 3	-.64***	-.49***	-.32*	-.52***	-.34*
Page 8	-.51***	-.53***	-.68***	-.69***	-.53***
Page 5	-.71***	-.77***	-.74***	-.77***	-.53***
Page 7	-.60***	-.72***	-.61***	-.60***	-.50***

* p< .05 level, **p<.01, *** p<.001 level (all values 2-tailed)

Jian Scale Trust-Oriented Items

	confident	security	integrity	dependable	reliable
Page 4	.55***	.72***	.69***	.63***	.77***
Page 6	.42**	.18	.41**	.76**	.47**
Page 2	.68***	.62***	.77***	.72***	.88***
Page 1	.76***	.64***	.74***	.74***	.89***
Page 3	.39**	.64***	.64***	.72***	.54***
Page 8	.79***	.48***	.76***	.81***	.56***
Page 5	.78***	.65***	.71***	.76***	.69***
Page 7	.76***	.61***	.61***	.76***	.71***

* p< .05 level, **p<.01, *** p<.001 level (all values 2-tailed)

Appendix P

Definitions for the Ten Colours in Study 3

(names and definitions are from http://www._anthus.com/Colors/NBS.html)

	NBS/IBC name	RGB decimal	RGB hex	Munsell scale Hue Value Chroma
Blue	Strong Blue	000 083 138	00 53 8A	2.9pb 4.1 10.4
Green	Strong Green	000 107 060	00 6B 3C	5.8g 4.4 8.7
Red	Vivid Red	193 000 032	C1 00 20	5.0r 3.9 15.4
Purple	Strong Purple	128 062 117	80 3E 75	6.5p 4.3 9.2
Yellow	Brilliant Greenish Yellow	255 220 051	FF DC 33	9.8y 8.8 9.5
Pastel Blue	Very Light Blue	166 189 215	6C 92 AF	2.7pb 7.9 6.0
Pastel Green	Very Light Green	152 199 147	98 C7 93	6.5g 7.8 4.9
Pastel Red	Vivid Pink	255 126 147	FF 7E 93	1r 8.0 13.0
Pastel Purple	Light Purple	186 127 162	BA 7F A2	6.2p 6.5 6.5
Pastel Yellow	No name	255 255 160	FF FF A0	Not available

Appendix Q

20-item Scale for Study 3

1. This Web site is deceptive.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

2. There is something underhanded about this Web site.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

3. I am suspicious of the intent of this Web site.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

4. I am wary of this Web site.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

5. If I used this Web site, it would have a harmful or injurious outcome.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

6. I am confident in this Web site.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

7. This Web site provides security.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

8. This Web site has integrity.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

9. This Web site is dependable.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

10. This Web site is reliable.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

Appendix Q (continued)

11. I can trust this Web site.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

12. This is an aesthetically pleasing Web site.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

13. This Web site is pleasant.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

14. This Web site is clear.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

15. This Web site is clean.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

16. This Web site is symmetric.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

17. This Web site is creative.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

18. This Web site is fascinating.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

19. This Web site is original.

Agree 1.....2.....3.....4.....5.....6.....7 Disagree

20. This Web site is sophisticated.

Disagree 1.....2.....3.....4.....5.....6.....7 Agree

Appendix R

RCBS Shyness Scale (Cheek, J.M., 1983).

Thank you for taking part in this study. The next step is to fill out this short questionnaire.

INSTRUCTIONS: Please read each item carefully and decide to what extent it is characteristic of your feelings and behaviour. Fill in the blank next to each item by choosing a number from the scale printed below.

- 1 = Very uncharacteristic or untrue, strongly disagree
- 2 = Uncharacteristic
- 3 = Neutral
- 4 = Characteristic
- 5 = Very characteristic or true, strongly agree

- ___ 1. I feel tense when I'm with people I don't know well.
- ___ 2. I am socially somewhat awkward.
- ___ 3. I do not find it difficult to ask other people for information.
- ___ 4. I am often uncomfortable at parties and other social functions.
- ___ 5. When in a group of people, I have trouble thinking of the right things to talk about.
- ___ 6. It does not take me long to overcome my shyness in new situations.
- ___ 7. It is hard for me to act natural when I am meeting new people.
- ___ 8. I feel nervous when speaking to someone in authority.
- ___ 9. I have no doubts about my social competence.
- ___ 10. I have trouble looking someone right in the eye.
- ___ 11. I feel inhibited in social situations.
- ___ 12. I do not find it hard to talk to strangers.
- ___ 13. I am more shy with members of the opposite sex.

Web Trustworthiness and Shyness

Appendix S

Study 3
n=84
t-tests Jian Trust Scale

Legend	t-value
	p-value

Jian	01	02	08	07	03	04	09	05	10	06
01 Blue		1.338 .185	1.464 .147	1.557 .123	2.522 .014	2.841 .006	2.605 .011	4.911 .000**	5.169 .000**	8.803 .000**
02 Green			0.210 .834	0.388 .699	1.063 .291	1.219 .226	1.539 .128	2.820 .006	3.241 .002	8.364 .000**
08 Purple				0.296 .768	0.823 .413	0.925 .357	1.858 0.067	2.777 .007	3.442 .001**	8.632 .000**
07 Pink					0.508 .613	0.521 .604	1.402 .164	2.410 .018	2.991 .004	8.523 .000**
03 Pastel Blue						0.097 .923	0.798 .427	2.648 .010	2.473 .015	9.206 .000**
04 Pastel Green							0.655 .514	2.222 .029	2.384 .019	7.885 .000**
09 Red								1.051 .297	1.361 .177	6.528 .000**
05 Pastel Purple									0.318 .751	7.447 .000**
10 Yellow										6.362 .000**
06 Pastel Yellow										

** significant at an overall experiment level $p < 0.05$ (Ryan, 1960)

Web Trustworthiness and Shyness

Appendix T

Jian Trust Scale

Study 3

n=84

Wilcoxon tests

Legend	Wilcoxon Z p-value
--------	-----------------------

Jian	01	08	07	02	03	04	09	10	05	06
01 Blue		1.562 .118	1.060 .289	1.509 .131	2.009 .045	2.583 .010	2.562 .010	4.593 .000**	4.259 .000**	7.072 .000**
08 Purple			0.223 .824	0.216 .829	1.031 .303	0.974 .330	1.757 .079	2.984 .003	2.732 .006	7.085 .000**
07 Pink				0.165 .869	1.068 .286	1.216 .224	1.824 .068	3.060 .002	3.078 .002	7.297 .000**
02 Green					1.157 .247	0.691 .490	0.699 .484	2.759 .006	2.712 .007	6.761 .000**
03 Pastel Blue						0.466 .655	0.622 .534	1.988 .047	1.752 .080	7.216 .000**
04 Pastel Green							0.29 .819	2.057 .040	2.016 .044	6.743 .000**
09 Red								1.532 .126	1.228 .220	6.006 .000**
10 Yellow									0.243 .808	5.852 .000**
05 Pastel Purple										6.341 .000**
06 Pastel Yellow										

** significant at an overall experiment level $p < 0.05$ (Ryan, 1960)

Appendix U

Study 3
 n=84 t-tests
 Visual Appeal Scale

Legend	t-value
	p-value

Jian	01	07	02	03	08	04	09	05	10	06
01 Blue							1.798 .076	3.675 .000**	5.906 .000**	11.336 .000**
07 Pink							1.297 .198	2.820 .006	4.128 .000**	11.264 .000**
02 Green								2.269 .026	3.973 .000**	11.331 .000**
03 Pastel Blue							1.330 .187	3.503 .001**	5.126 .000**	12.232 .000**
08 Purple							1.169 .246	2.318 .023	4.112 .000**	11.532 .000**
04 Pastel Green								2.176 .032	3.981 .000**	11.077 .000**
09 Red								1.008 .316	2.860 .005	9.340 .000**
05 Pastel Purple									2.526 .013	10.437 .000**
10 Yellow										7.803 .000**
06 Pastel Yellow										

** significant at an overall experiment level $p < 0.05$ (Ryan, 1960)

Web Trustworthiness and Shyness

Appendix V
 Study 3
 n=84 t-tests
 Wilcoxon Results
 Visual Appeal

Legend	Wilcoxon Z p-value
--------	-----------------------

Visual Appeal	01	03	08	07	04	02	09	05	10	06
01 Blue						0.675 .500	1.329 0.184	3.111 0.002	5.112 0.000**	7.415 0.000**
03 Pastel Blue							1.608 0.108	3.062 0.002	4.812 0.000**	7.563 0.000**
08 Purple							0.860 0.390	2.232 0.026	3.987 0.000**	7.549 0.000**
07 Pink							1.261 0.207	7.963 0.000**	4.056 0.000**	7.644 0.000**
04 Pastel Green							0.225 0.822	2.116 0.034	3.525 0.000**	7.362 0.000**
02 Green							0.958 0.338	1.974 0.048	3.829 0.000**	7.442 0.000**
09 Red								1.169 0.242	3.133 0.002**	6.934 0.000**
05 Pastel Purple									2.684 0.007**	7.264 0.000**
10 Yellow										6.353 0.000**
06 Pastel Yellow										

** significant at an overall experiment level $p < 0.05$ (Ryan, 1960)

Appendix W

Study 3

Trust scores

Maximums, means, and minimums for Shy and Non-Shy groups

	01 Blue		02 Green		03 Pastel Blue		04 Pastel Green		05 Pastel Purple	
	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy
maximum	63	76	63	74	60	72	68	70	65	68
mean	43.9	56.4	45.1	52.3	44.3	49.2	44.1	50.2	42.6	48.6
median	43.5	58.0	44.0	55.5	44.0	49.5	43.5	50.5	44.0	50.0
minimum	25	25	26	17	30	17	30	22	21	14
	06 Pastel Yellow		07 Pink		08 Purple		09 Red		10 Yellow	
	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy
maximum	63	73	70	76	63	74	68	74	71	77
mean	34.1	37.1	42.5	55.1	44.5	52.1	43.9	50.3	41.4	49.7
median	31.5	36.0	40.5	55.5	42.5	53.0	41.5	48.5	39.5	50.0
minimum	21	14	26	16	24	19	23	17	20	17

Visual Appeal scores

Maximums, means, and minimums for Shy and Non-Shy groups

	01 Blue		02 Green		03 Pastel Blue		04 Pastel Green		05 Pastel Purple	
	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy
maximum	45	59	47	55	51	52	51	56	50	49
mean	32.4	38.9	34.0	36.6	35.1	36.4	33.9	37.6	32.0	34.8
median	30.0	39.5	34.5	35.0	34.5	36.0	33.5	39.0	32.0	35.0
minimum	24	17	22	18	24	15	19	17	17	15
	06 Pastel Yellow		07 Pink		08 Purple		09 Red		10 Yellow	
	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy	Shy	Non-shy
maximum	34	50	56	57	45	60	53	61	46	50
mean	22.3	21.8	33.1	39.7	32.6	37.1	34.3	33.5	29.8	33.7
median	22.5	20.0	31.5	40.0	32.0	37.5	34.5	32.0	29.0	32.0
minimum	10	10	21	9	21	14	20	9	16	12

Appendix X

Study 3 correlations for each page within Shy and Non-shy groups

Shy	Blue	Green	Pastel Blue	Pastel Green	Pastel Purple	Pastel Yellow	Pink	Purple	Red8	Yellow
Pearson	.677	.566	.612	.708	.622	.681	.696	.806	.627	.682
p value	.001	.009	.004	.000	.003	.001	.001	.000	.003	.001
Spearman	.620	.471	.642	.628	.587	.689	.521	.680	.468	.543
p value	.004	.036	.002	.003	.007	.001	.018	.001	.038	.013

Non-shy	Blue	Green	Pastel Blue	Pastel Green	Pastel Purple	Pastel Yellow	Pink	Purple	Red8	Yellow
Pearson	.780	.600	.646	.632	.602	.707	.734	.699	.844	.630
p-value	.000	.005	.002	.003	.005	.000	.000	.001	.000	.003
Spearman	.698	.541	.565	.546	.483	.537	.618	.674	.848	.519
p-value	.001	.014	.009	.013	.031	.015	.004	.001	.000	.019

Appendix Y

Study 4. 27 Photos

Couples

<p>Couple #1</p> 	<p>Couple #2</p> 	<p>Couple #3</p> 
<p>Couple #4</p> 	<p>Couple #5</p> 	<p>Couple #6</p> 
<p>Couple #7</p> 	<p>Couple #8</p> 	<p>Couple #9</p> 

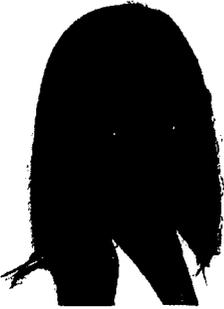
Appendix Y (continued)

Men

Man #1	Man #2	Man #3
 A black and white portrait of a man with short dark hair, wearing a dark shirt and a light-colored tie. He is looking slightly to the left of the camera with a neutral expression.	 A black and white portrait of a man with short dark hair, wearing a dark shirt and a light-colored tie. He is looking directly at the camera with a neutral expression.	 A black and white portrait of a man with short dark hair, wearing a light-colored shirt and a dark tie. He is smiling broadly at the camera.
Man #4	Man #5	Man #6
 A black and white portrait of a man with short dark hair, wearing a dark shirt. He is smiling at the camera.	 A black and white portrait of a man with short dark hair, wearing a dark shirt and a light-colored tie. He is smiling at the camera.	 A black and white portrait of a man with short dark hair and glasses, wearing a light-colored shirt and a dark tie. He is smiling at the camera.
Man #7	Man #8	Man #9
 A black and white portrait of a man with short dark hair, wearing a dark shirt. He is smiling at the camera.	 A black and white portrait of a man with short dark hair and glasses, wearing a dark shirt. He is smiling at the camera.	 A black and white portrait of a man with short dark hair, wearing a dark shirt and a light-colored tie. He is smiling broadly at the camera.

Appendix Y (continued)

Women

Woman #1	Woman #2	Woman #3
 A black and white portrait of a woman with dark hair pulled back, wearing a dark top and a necklace. She is smiling slightly.	 A black and white portrait of a woman with dark hair, wearing a dark top and a necklace. She is smiling.	 A black and white portrait of a woman with dark hair, wearing a dark top and a necklace. She is smiling.
Woman #4	Woman #5	Woman #6
 A black and white portrait of a woman with dark hair, wearing a dark top and a necklace. She is smiling.	 A black and white portrait of a woman with long dark hair, wearing a dark top. She is smiling.	 A black and white portrait of a woman with dark hair, wearing a dark top and a necklace. She is smiling.
Woman #7	Woman #8	Woman #9
 A black and white portrait of a woman with long dark hair, wearing a dark top. She is smiling.	 A black and white portrait of a woman with dark hair, wearing a dark top and a necklace. She is smiling.	 A black and white portrait of a woman with long dark hair, wearing a dark top and a necklace. She is smiling.

Appendix Z

The Preparation of the “friendliness” Scale for Study 4

A semantic-differential scale was prepared as follows. From a search at <http://thesaurus.com> using the search term <friendly> I found 45 synonyms and nine antonyms. As well, from fellow teachers, I gleaned four other words that were thought to relate to “friendly”. Altogether this generated 59 words: the word itself, 49 synonyms and nine antonyms. In conversations with fellow teachers, I changed “not cold” to “warm” and “not unreceptive” to “receptive”; I deleted 10 words including “not unfriendly”, “not unsociable”, and “not cool”. This left 49 words to be rated: affable, affectionate, amiable, amicable, attached, attentive, auspicious, beneficial, benevolent, benign, chummy, civil, close, comradely, conciliatory, confiding, convivial, cordial, faithful, familiar, favourable, fond, friendly, genial, good, helpful, kind, kindly, loving, loyal, neighbourly, on good terms, outgoing, peaceable, peaceful, propitious, receptive, sociable, solicitous, sympathetic, tender, welcoming, well-disposed, warm, receptive, comforting, happy, approachable, likeable.

Seven university-trained former teachers rated each of the 49-items on a 7-point semantic-differential scale, ranging from “NONE” to “100%” as per the instruction to: “Rate these words on how well they are related to describing a “friendly face”.

Words which were rated as “6” or “7” by four or more raters were accepted, providing no rater had given the word a “0” This yielded 8 words: affable, friendly, genial, happy, kindly, sympathetic, warm, and welcoming.

Appendix Z (continued)

	Words tested	Scores							
		0	1	2	3	4	5	6	7
	friendly						2	1	4
	affable					1	2	1	3
	genial					2	1	1	3
	kindly				1	1	1	2	2
	sympathetic					1	2	2	2
	Welcoming					1	2	2	2
	Warm				1	1	1	2	2
	Happy					1	2	2	2

This allowed for the construction of an 8-item, 7-point semantic differential scale (Appendix AA). The scale items were balanced left to right, with four of the items having the “agree” anchor to the left and four to the right, to control for any tendency to respond in a particular direction. Four items were of the form “This is a [.....] face; four of the form “This face is [.....]”.

Appendix AA

The "friendliness" Scale Study 4

1.	This is a friendly face.
Agree	1.....2.....3.....4.....5.....6.....7 Disagree
2.	This face is affable.
Agree	1.....2.....3.....4.....5.....6.....7 Disagree
3.	This face is genial.
Disagree	1.....2.....3.....4.....5.....6.....7 Agree
4.	This is a kindly face.
Agree	1.....2.....3.....4.....5.....6.....7 Disagree
5.	This is a sympathetic face.
Disagree	1.....2.....3.....4.....5.....6.....7 Agree
6.	This face is welcoming.
Agree	1.....2.....3.....4.....5.....6.....7 Disagree
7.	This face is warm.
Disagree	1.....2.....3.....4.....5.....6.....7 Agree
8.	This is a happy face.
Disagree	1.....2.....3.....4.....5.....6.....7 Agree

Appendix AB

Pearson Product Moment Correlations
for items in the "friendliness" scale
for Study 4

	Q#2	Q#3	Q#4	Q#5	Q#6	Q#7	Q#8	Total 8 Qs
Q#1 friendly	.642***	.497***	.784***	.230***	.734***	.608***	.605***	.835***
Q#2 affable		.472***	.669***	.292***	.679***	.442***	.456***	.793***
Q#3 genial			.499***	.471***	.471***	.663***	.598***	.715***
Q#4 kindly				.241***	.698***	.581***	.540***	.822**
Q#5 sympathetic					.259***	.430***	.288***	.498***
Q#6 welcoming						.549***	.549***	.830***
Q#7 warm							.675***	.760***
Q#8 happy								.713***
Total 8 Qs								

*** significant at p<.001

Web Trustworthiness and Shyness

Appendix AC

Study 4 Friedman Ranks and Significant Wilcoxon Tests

Men	Friedman rank	#4	#3	#5	#9	#8	#6	#2	#1	#7
#4	7.23			2.16*	2.56*	3.44**	3.19**	3.03**	3.70***	3.75***
#3	6.33					2.54*	2.38*		3.12**	3.14**
#5	6.00					2.06*		1.99*	3.27**	3.31**
#9	5.53								3.01**	3.31**
#8	4.78								2.87**	2.30*
#6	4.68								2.15*	2.30*
#2	4.63								2.25*	2.50*
#1	2.95									
#7	2.90									

Women	Friedman rank	#5	#6	#3	#2	#9	#8	#1	#4	#7
#5	6.425						2.23*		2.94**	3.79***
#6	5.925						1.99*		2.20*	3.34**
#3	5.75									2.99**
#2	5.15									2.82**
#9	5.10									2.92**
#8	4.825									2.12*
#1	4.725									2.44*
#4	4.40									
#7	2.70									

Couples	Friedman rank	#9	#7	#1	#2	#4	#5	#6	#3	#8
#9	6.20							2.29*	2.54*	3.72***
#7	6.20								2.62**	3.62***
#1	5.88								2.80**	3.76***
#2	5.73								2.25*	3.83***
#4	5.20									3.04**
#5	5.03									2.92**
#6	4.70									2.96**
#3	3.98									3.44**
#8	2.10									

* sig. at $p < .05$, ** sig. at $p < .01$, *** sig. at $p < .001$ (all 2-tailed)

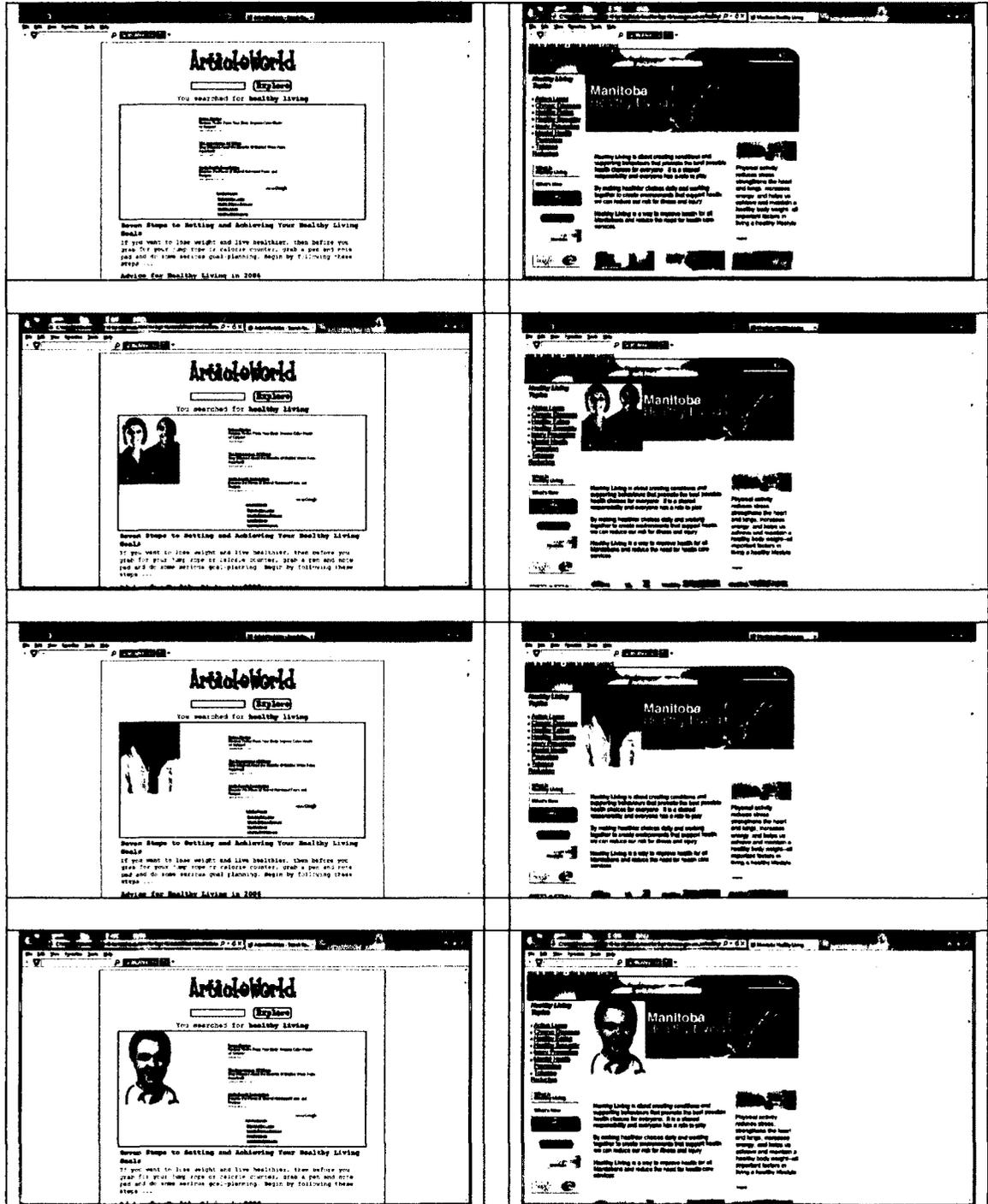
Appendix AD

Photos for Study 5

Couple #9	Couple #7
	
Woman #5	Woman #6
	
Man #4	Man #3
	

Appendix AE

14 Web Pages for Study 5



Appendix AF

Study 5. Wilcoxon Results for Seven Pages

ArticleWorld pages

Jian trust	Friedman rank	Wilcoxon comparison to plain ArticleWorld page (AP)
AW#5	4.53	Z=2.380 p=.017*
AC#9	4.47	Z=2.290 p=.022*
AM#4	4.11	Z=1.688 p=.091
AC#7	3.92	Z=1.253 p=.210
AW#6	3.84	Z=0.918 p=.359
AM#3	3.78	Z=1.159 p=.246
AP	3.34	

* sig. at $\alpha < .05$ (2-tailed)

Manitoba pages

Jian trust	Friedman rank	Wilcoxon comparison to plain Manitoba page (MP)
MP	4.34	Z=2.380 p=.017*
MW#5	4.17	Z=0.991 p=.322*
MC#9	4.11	Z=1.371 p=.170
MM#4	4.07	Z=1.331 p=.183
MW#6	3.95	Z=1.603 p=.109
MC#7	3.77	Z=1.217 p=.224
MM#3	3.57	Z=2.414 p=.016

* sig. at $\alpha < .05$ (2-tailed)

Appendix AG

Results of Mann-Whitney Test
 Study 5
 Shy versus Non-Shy Trust ranks

	ArticleWorld sites		Manitoba sites	
	HiShy mean rank	LoShy mean rank	HiShy mean rank	LoShy mean rank
Plain	23.95	17.05	16.73	24.28
Couple #7	24.60	16.40	18.55	22.45
Couple #9	23.68	17.33	17.55	23.45
Man #3	22.88	18.13	17.30	23.70
Man #4	22.35	18.65	18.45	22.55
Woman #5	22.70	18.30	18.48	22.53
Woman #6	20.70	20.30	18.18	22.83

Appendix AH

Median Scores for 14 Web Pages Study 5

Median scores		
Non-shy group		Shy group
Manitoba w/Woman #5	69	
Manitoba plain	67	
Manitoba w/Man #3	66	
Manitoba w/Couple #7	65.5	
Manitoba w/Woman #6	64.5	
Manitoba w/Man #4	64	
Manitoba w/Couple #9	63.5	
	61.5	Manitoba w/Woman #6
	60.5	Manitoba w/Couple #9
	59.5	Manitoba plain
	59	Manitoba w/Woman #5
	58	Manitoba w/Couple #7
	57.5	Manitoba w/Man #4
	52.5	Manitoba w/Man #3
	33.5	ArticleWorld w/Woman #5
	33	ArticleWorld w/Man #4
	33	ArticleWorld w/Woman #6
	33	ArticleWorld w/Couple #9
	32	ArticleWorld w/Couple #7
	31	ArticleWorld w/Man #3
	30	ArticleWorld plain
ArticleWorld w/Man #4	29	
ArticleWorld w/Woman #6	28.5	
ArticleWorld w/Woman #5	28.5	
ArticleWorld w/Couple #7	27	
ArticleWorld w/Man #3	26.5	
ArticleWorld w/Couple #9	24	
ArticleWorld plain	23	