Preprint version. Final published paper: doi:10.1145/3533047

What is Beautiful is Secure

MILICA STOJMENOVIĆ, Carleton University, Canada
ERIC SPERO, Carleton University, Canada
MILOŠ STOJMENOVIĆ, Singidunum University, Serbia
ROBERT BIDDLE, Carleton University, Canada

Visual appeal has been shown to influence perceptions of usability and credibility, and we hypothesize that something similar is happening with user judgments of website security: *What is beautiful is secure*. Web certificates provide reliable information about a website's level of security, presented in browser interfaces. Users should use this to inform their trust decisions online, but evidence from laboratory studies and real-world usage suggests that they do not. We conducted two studies—one in lab, and one online—in which participants view and interact with websites with high and low visual appeal, and various security levels, and then make security-related judgments. In both studies, participants consistently rated visually appealing websites as more secure, and indicated they would be more likely to enter sensitive information into visually appealing websites—even when they were less secure. Our results provide evidence that users rely on visual appeal when making security and trust decisions on websites. We discuss how these results may be used to help users.

CCS Concepts: • **Security and privacy** → *Usability in security and privacy*; • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: aesthetics, visual appeal, cybersecurity, web certificates, perceived usability, human-computer interaction

ACM Reference Format:
Milica Stojmenović, Eric Spero, Miloš Stojmenović, and Robert Biddle. 2022. What is Beautiful is Secure. *ACM Trans. Priv. Sec.* 1, 1, Article 1 (January 2022), 30 pages. https://doi.org/10.1145/3533047

1 Introduction

It has been shown that *What is beautiful is usable* [53], and we speculate that what is beautiful is *secure* as well. The influence of a perceived attribute of an object on an unrelated other attribute is known in the psychology literature as a 'halo effect'. In the landmark paper that first demonstrated this effect, *What is beautiful is good* [10], people rated as beautiful were perceived to have more socially desirable traits. The halo effect was later observed in the context of HCI and user interfaces. For example, users are quick to form a number of judgments about a website based on their first impression of its appearance [23]. In the domain of cybersecurity, it has been shown that users in general do not pay

Authors' addresses: Milica Stojmenović, milica.stojmenovic@carleton.ca, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6; Eric Spero, eric.spero@carleton.ca, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6; Miloš Stojmenović, mstojmenovic@singidunum.ac.rs, Singidunum University, Danijelova 32, Belgrade, Vojvodina, Serbia, 160622; Robert Biddle, robert.biddle@carleton.ca, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6.

attention to security information in browser interfaces. Remarks made by participants in a study on alternative certificate interfaces suggest that users were making judgments about a website's identity based on its visual appearance [48].

Given that many activities relating to both work and play are today carried out online, online security and best practices become critical. This becomes more important when we realise that users make many key decisions about which websites to trust, which ones to make purchases from, and which ones to log into (revealing passwords and personal details). What do users rely on to make security decisions online? We specifically wish to study the case of websites that are fraudulent, but mimic trustworthy sites to capture credentials, violate users' security and privacy, download malware, and undermine infrastructure. This paper examines if people base their judgments of website security on visual appeal.

Aesthetics has been defined in many ways [21]. However, the two main definitions are: (1) to describe a pleasant physiological reaction or (2) a visual property of an object requiring judgment of its appearance [11]. Henceforth, we refer to the latter definition as visual appeal. We take the following definition of cybersecurity, which is meant to encompass its many dimensions: "cybersecurity is the organization and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign de jure from de facto property rights" [9]. Usability is defined by the International Standards Organization [16] as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". This is the definition used in this paper as well.

Website security is often technically defined as the support for encrypted connections (e.g. TLS). We are primarily concerned with a broader, more commonsense notion of security which includes, in addition to the technical definition, an indication of the website's identity, which can be used to inform user trust decisions. This information could help users guard against a subset of attacks—any that attempt to mislead users about a website's identity, such as phishing attacks.

The primary objective of this paper is to examine the influence of website visual appeal on perceived security. We want to know if what is beautiful is regarded as *secure*. As a secondary goal, we aim to replicate previous research demonstrating an influence of visual appeal on usability.

2 Background

2.1 The Halo Effect and the Impact of Visual Appeal

In psychology, it has been shown that perceptions of object attributes can affect perceptions of other, unrelated attributes. For example, people may make character assumptions based on a person's appearance. This is called the halo effect, where, for example, beautiful people are perceived to have more socially desirable traits [10]. In HCI, in spite of some methodological issues [15], it has been demonstrated that visual appeal influences a user's first impression of a website [53], and that this first impression often happens in under 50 ms [24]. Research in HCI has also shown that this first impression can impact perceptions of other attributes. We present a summary of a small sample of the most central papers on this topic here.

While research in the area existed [20] prior to the work of Tractinsky and colleagues [52, 53], they seem to have catalyzed a stream of research on the influence of visual appeal on usability in HCI. They found that the visual appeal of ATM interfaces affected usability, where more visually appealing interfaces were rated more usable, and that visually appealing interfaces increased satisfaction as well as perceptions of quality, both before and after use [53]. A few years later, [33] presented a framework for emotional design in which the visual appearance of an object has a deep influence

on how it is judged by the user. Many studies followed Tractinsky and colleagues, examining the relationship between visual appeal and usability. For example, [54] found that usability is affected by visual appeal before the system is even used. [49] found a complementary relationship between visual appeal and usability: visual appeal impacts user experience by creating a captivating first impression, whereas usability is more important to keep users engaged with a website during and after use. Two other experiments by [56] showed that there was a preference for relatively attractive pages over relatively unattractive pages after short exposures, and aesthetic pages that are also more informative are rated as more attractive than purely visually appealing pages.

Visual appeal has also been found to influence other factors in addition to usability. For example, [38] show that visual appeal influences judgments about the credibility of a website's content, and the effect registers within the first few seconds of looking at the page. Perceptions of visual appeal affect user preferences before using the website [22]. In addition, [23] measured subjective ratings of visual appeal, usability, and trustworthiness after showing websites for a short time (500ms), and found evidence that all three ratings were driven by visual appeal. Other studies in HCI [27, 28, 51] have also found ties between visual appeal and emotion, and both can influence perception of websites.

The halo effect of visual appeal has also been observed in domains outside HCI. For example, researchers have linked visual appeal with measures of health and fertility [14, 26, 37, 43], suggesting that the halo effect may have emerged as a consequence of evolutionary processes. Visual appeal also appears to affect food choice. When selecting which food item to purchase, customers are influenced by visual elements such as colour, graphics, size, and shape [42]. Silayoi and Speece argue that this is because visual elements are easier to process than informational elements, and visual elements evoke a stronger emotional response. Another study in the domain of food choice showed that food products with aesthetic packaging were chosen more often—even at higher prices, and over well-known brands; that these decisions were slower on average, indicating a stronger affective response; and exposure to these aesthetic items increased activation in a brain area associated with reward [36]. In the context of industrial product choice, [61] found that the visual appeal of industrial products had an influence on preference that often exceeded the influence of performance factors and cost. These and other previous findings mentioned here motivated us to examine the impact of visual appeal on perceived security.

2.2 Online Security

Web certificates offer two main assurances to users. The first is that the connection between the browser and the website is encrypted. Encryption provides a secure channel of communication between users and websites. Secure Sockets Layer (SSL) and Transport Layer Security (TLS) use asymmetric cryptography for in-transit encryption.

Although a connection between a user and a website may be 'secure' strictly speaking through encryption, if the secure connection is to a malicious website this is obviously undesirable. The second assurance offered by certificates pertains to information about the website's identity. Certificate Authorities (CAs) create and issue X.509 certificates [8] to website owners; depending on the type of certificate requested, CAs will take steps to verify the identity of the certificate owner. Website identity information can help users in determining whether the website indeed belongs to an identity they trust.

There are three types of certificates accepted by web browsers: Domain Validated (DV), Organization Validated (OV), and Extended Validation (EV). All three certificates ensure that the connection is encrypted. DV confirms that the certificate is issued to someone who controls the domain, determined by a challenge-response process. OV adds details of the website's organization (e.g. company registration), which is confirmed by CAs. EV adds more identification

information, including the geographic location of the organization, confirming contact details, and linking the website to a legal entity. The majority of legitimate websites have OV certificates [45].

Web servers store the certificates, and they are accessed by browsers. When a user accesses a website with the `https` protocol, the browser first retrieves the certificate. If the CA that issued the certificate is recognized and the certificate is valid, then the browser uses it to establish encrypted access to the website, and presents identity information to the user, depending on the type of certificate. A quick summary of certificate information is shown in the certificate indicator, to the left of the URL. More detailed information is available in a menu that is accessible by clicking the certificate indicator. However, this information is not only difficult to find, but uses obscure language that is meaningless to all but the most technical users.

The three validation types all provide the same assurances regarding connection security, but they differ significantly in assurances of *identity*. DV certificates provide next to no identity assurances, as they are easily obtainable by anyone with a registered domain. Shopify, a service that helps vendors create websites, promoted the fact that it provides free DVs for all sites [41]. Most relevant to the present study, even malicious actors can easily obtain DV certificates for fraudulent websites from services like Let's Encrypt; in late 2020 it was estimated that 89% of phishing sites were Domain Validated [1], and this figure rises each year. To acquire an EV certificate, on the other hand, the organization belonging to the certificate holders has been vetted to some extent, and they can in principle be held legally responsible if the user is harmed while using the site.

Currently (mid-2022) the certificate indicators used by Google Chrome[1] are as shown in Figs. 1b–1e. EV, OV, and DV certificates are all given a lock symbol. Prior to this, until mid-2019, the certificate indicator for EV had additional text denoting its identity and the company's national registration jurisdiction. This 'legacy' EV is shown in Fig. 1a. In announcing their decision to remove the EV indicator, Google cited that the indicators did not appear to dissuade users from making insecure choices, and the company name information provided could be confusing to users [6].

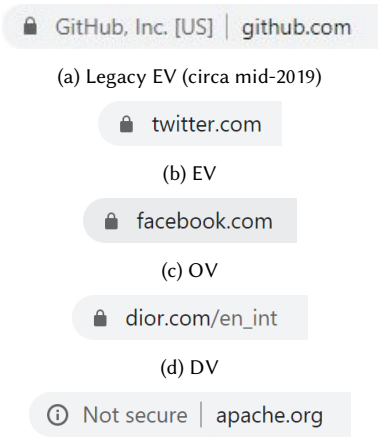
EV certificates today (Fig. 1b) are not differentiated from OV or DV in the browser UI. In these examples, Twitter has an EV certificate [2] and Facebook has an OV; both offer additional identity information compared to Dior's DV. Yet all three indicators appear the same, masking from users the additional assurances offered by OV and EV.

This lack of differentiation between the various certificate types in browser UIs means that, dangerously, a fraudulent website with a DV would appear to end users the same as a legitimate website featuring an OV or EV certificate: all featuring the lock symbol, which has strong connotations of security. For example, secwww.com/facebook (owned by us) shows a valid DV, and could easily be made to look identical to the real Facebook website thanks to freely-available website download tools.

Assurances of identity provided by web certificates can help users guard against attacks where there is an attempt to mislead users about who a site belongs to. Common among these are phishing attacks in which a malicious site is made to look and function similar to a legitimate site, and with a similar-looking domain name. This is the type of attack we focus on in this paper. In other attacks, such as when a legitimate website is hijacked or affected by malware, certificates will not be able to help.

While web browsers have removed the indicator for EV certificates, EV certs can still have an important role in user safety by helping users distinguish fraudulent from genuine websites. As mentioned earlier, the overwhelming majority of phishing websites feature Domain Validated certificates (offering next to no assurances about identity), appearing as 'secure' (via the lock symbol) to the end user. Due to the short-lived nature of phishing websites—most phishing sites

[1]Throughout this paper we focus on Google Chrome, as it is by far the dominant web browser in terms of usage [30, 58].
[2]Twitter no longer uses an EV certificate.



(a) Legacy EV (circa mid-2019)

(b) EV

(c) OV

(d) DV

(e) No certificate

Fig. 1. Certificate indicators and URLs as shown in Google Chrome, mid-2019 and current (mid-2022). (Note: At the time of our original study Twitter used an EV certificate, but they have since switched to an OV.)

exist for less than 24 hours [2]—blocklist solutions such as Google Safe Browsing are not an effective solution. The assurances that come with EV and even OV certificates could be a key factor in helping users avoid phishing attacks.

In light of the potentially confusing browser certificate interfaces, the question is: What do users rely on to make their decisions about which websites are secure? Since at the time of this study the indicators for EV, OV, and DV were visually identical, our goal was to examine the impact of visual appeal on perceived security, in an effort to understand if this is an important factor in user decision-making.

2.3 Studies on Website Authentication and Certificate Interfaces

We now summarize the most relevant studies on website certificate interfaces. Earlier work [39, 60] has shown that users, in general, do not notice security information in browser UIs. [12] investigated alternative indicators before implementing them in the Google Chrome browser, but their indicators did not address identity. [48] tested a browser extension that displayed popups showing a website's security status to users, and found that these were largely ignored as well—although they were more effective when placed in the upper-right of the window. In post-task interviews with participants, some reported making decisions about a website's identity based on its visual appearance. Additionally, a paper by [3] mentioned that participants in their study seemed to have been relying on appearance when rating security.

These findings, taken together with the 'halo effect' of visual appeal, suggest that some users were perhaps making online security decisions on visual appeal. In the absence of clear indicators, we speculated that people will make judgments on security based on visual appeal since it was found to be an influential factor in other HCI domains, such as usability, outlined in the next section.

3 Dataset

3.1 Existing Website Database

In order to experimentally study the effect of visual appeal on perceived security, we needed to control for certain factors: Mainly visual appeal and security, but we also wanted to eliminate potential confounding variables, such as usability and prior experience. For visual appeal, we needed a website that was very appealing and also one that was not appealing to users. However, having two different websites to fill these roles would introduce confounding factors, including potentially different usability levels. Therefore, we needed to have a single website that could be manipulated to fit the required visual appeal levels. In everyday life, broken things can be perceived as less secure (e.g. a broken lock, window, or door). It was therefore important to ensure that usability was held constant across the high and low visual appeal websites to eliminate the possibility of usability influencing ratings of visual appeal or security.

Our dataset is based on the Gold Coast City Council website (Fig. 2a) and a low visual appeal variant of this website (Fig. 2b). The Gold Coast website was chosen for its user-perceived characteristics as determined by prior studies [46]. This website was selected among 52 websites as the most visually appealing, and user and expert-based usability testing showed that it was reasonably easy to use. It was also found to be unfamiliar to users, which is needed to eliminate potential confounds of prior experience. The websites were all in English.

In the Lva variant of the Gold Coast website, the colours of the original images were inverted, and the background colours were changed from beige to lilac, and from grey to evergreen [46]. The colour of the textual background (white) was not changed to help preserve perceived usability. These manipulations were found to be effective [46]: users rated the original website as significantly more visually appealing than the low visual appeal variant. Importantly, the manipulations preserved usability, as there was no significant differences in perceived usability of the two websites [44].

3.2 Additional Changes and Website Conditions

The high visual appeal (Hva) and low visual appeal (Lva) websites allow us to examine the impact of visual appeal on judgments of website security. We made further manipulations to the indicated security levels of the websites, through their indicated certificates and URLs, to help provide context to any findings of effects relating to visual appeal.

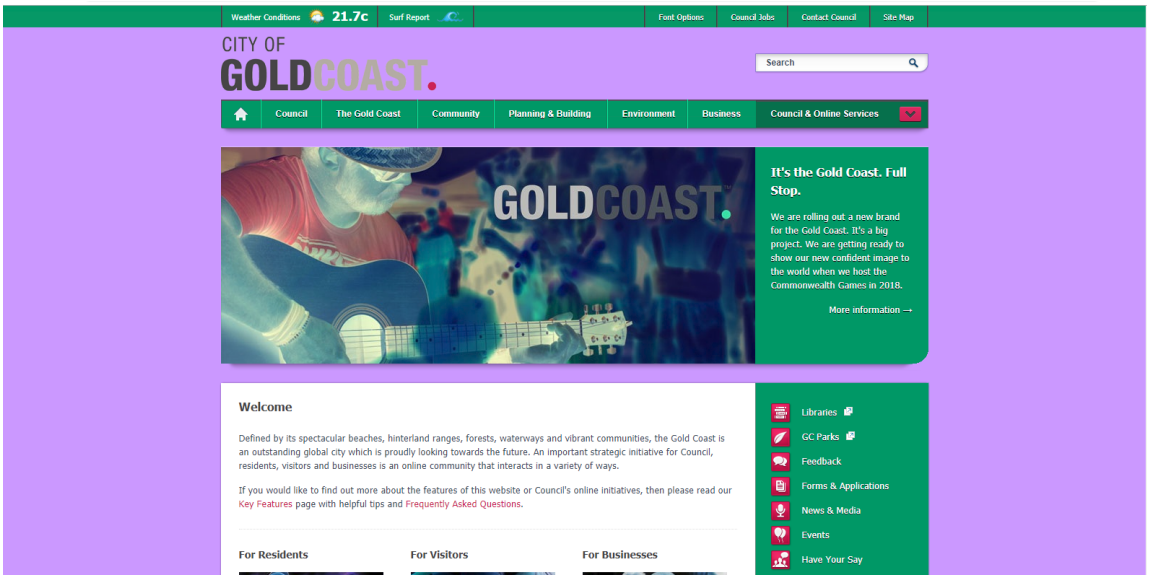
There were four security/certificate levels: EV (3), OV (2), DV (1), and no certificate (0), and two URL levels: 'real' and 'fake'. The 'real' URL matched the actual URL used by the Gold Coast City Council website (goldcoast.qld.gov.au). The 'fake' URL (citygoldcoast.com; owned by us) was chosen because it was available for purchase by anyone, and could conceivably pass as the Gold Coast's actual URL. We think it therefore approximates what might be used in an attack site impersonating the Gold Coast City Council.

The 'real' URL was paired with the EV and OV certificates, and the 'fake' URL with the DV certificate and no-certificate condition. These pairings were also chosen for their real-life plausibility: it is easy for a fraudulent website to acquire a DV certificate, but not OV or EV.

The URL-certificate pairings seen by participants are shown in Fig. 3. These pairs were combined once each with the high and low visual appeal websites in Fig. 2, giving a total of eight conditions. These conditions, shown in Table 1, enable us to separately compare the effects of visual appeal and indicated security on user perceptions of security.

All websites were hosted locally on our lab's servers, and were only accessible from inside our lab—including our copy of the Gold Coast website bearing its 'real' URL.

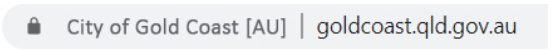
(a) The high visual appeal (HVA) website.



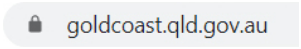
(b) The low visual appeal (LVA) website.

Fig. 2. The websites used in Study 1 and Study 2.

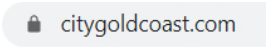
In Study 1 (Section 5), the conditions used were L2 and H1. In Study 2 (Section 6), all eight conditions were used. To reduce confounds such as learning effects and boredom, the eight conditions were split into two groups, so each participant only saw four conditions.



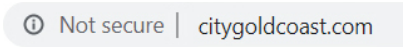
(a) Legacy EV (circa mid-2019), 'real' URL



(b) OV, 'real' URL



(c) DV, 'fake' URL



(d) No certificate, 'fake' URL

Fig. 3. Certificate indicators and URLs shown to participants (paired with websites in Fig. 2) in Study 2

Table 1. Conditions used across the two studies in this paper.

	Visual Appeal Level	
Security Level	High (H)	Low (L)
EV: 3	H3	L3
OV: 2	H2	L2
DV: 1	H1	L1
No certificate: 0	H0	L0

4 Research Questions and Hypotheses

We used these websites in studies with users where we sought an answer to the question: *Do users rely on visual appeal to make security judgments on websites?*

Our main goal was to test the hypothesis: What is beautiful is secure. Specifically, our hypothesis was that a highly appealing website would yield higher levels of perceived security than a visually unappealing website, regardless of what security indicator is shown. If people rated the more visually appealing website as more secure, regardless of its actual security level, then we could conclude that visual appeal drove user perceptions of security.

As a secondary goal, we wanted to replicate the findings of previous work which showed an influence of visual appeal on judgments of website usability. The second hypothesis is that participants will rate the more visually appealing websites as more usable, in spite of the fact that the websites' actual usability levels are identical.

We examine these research questions across two studies, one in-lab and one online. The two studies are described in Sections 5 and 6 below.

5 Study 1: In-Lab Study

The first study was carried out in-person, in our HCI research lab. Participants interacted with live websites, and completed questionnaires assessing their perceptions of visual appeal, security, and usability. In the context of security, having time to browse a website and interact with the browser's certificate indicator would give users the identity information they need to determine authenticity. Therefore, in Study 1 of this paper, we allowed users time to browse the

websites they were presented, which allowed them the opportunity to gather more information than first impressions alone could provide.

Our hypothesis was that a visually appealing website with weak security would yield higher security ratings from participants than a website that was lower in visual appeal but had higher security. If people rated the more visually appealing website as more secure, regardless of its actual security level, then we could conclude that visual appeal drove user perceptions of security. As a secondary goal, we wanted to replicate the findings of previous work which showed an influence of visual appeal on judgments of website usability. We predicted that participants would rate the more visually appealing website as more usable, even though the websites' actual usability levels were identical.

The aim of Study 1 was to test the two hypotheses in a simple experiment, examining two opposing cases: High visual appeal with DV and low visual appeal with OV (which provides website identity information in the certificate interface). These two certificates are only distinguishable by interaction, since they have the same certificate indicators. By holding as many factors constant as possible between the two conditions and minimizing confounds, the differences in participant ratings between the two websites could more readily be attributed to visual appeal. This initial study focused on opposing cases to quickly determine if there was evidence supporting the hypotheses.

Participants were able to interact with the live websites, allowing them the opportunity to explore both the website, and information pertaining to its certificate, displayed in the browser UI. The high visual appeal (HVA) website had a DV certificate, which offers no ownership or identity information, and a 'fake' URL. The low visual appeal (LVA) website had an OV certificate which provided the website's website identity information, and the website's 'real' URL. This study was cleared by our institute's research ethics board.

5.1 Method

5.1.1 Participants Twenty-one volunteers participated in this study (14 females, 7 males). Twelve were aged 18–19, eight were 20–29, and one was over 30. All 21 people were educated or working on their university degrees, but none were experts in cybersecurity. Nine had backgrounds in psychology, sociology, or cognitive science, four people had backgrounds in each of: computer science or engineering, biology/chemistry/medicine, and law/criminology/forensics. All participants had 20/20 or corrected to 20/20 vision, and screened for colour-blindness (to ensure that they could differentiate between the nuances in visual appeal). All participants were technology-savvy regular Internet users, stating that they used mobile phones, laptops and/or desktop computers on a daily basis. Sixteen used Google Chrome as their predominant browsers, four used Apple Safari, and one used Microsoft Internet Explorer/Edge. Participants were individually tested, with each session lasting approximately one hour. All participants were English-speakers.

5.1.2 Apparatus and Location Participants were tested using a PC computer, running Windows 10. The study took place in a private room inside a human-computer interaction laboratory. The participant and the researcher were in the same room, with the researcher interacting with the participant to explain task descriptions on the computer screen that the participants were interacting with, and to probe participants for further insights. Participants' audio and video were not recorded, but an eye-tracker (Gazepoint GP3 Desktop) was used to capture what participants were looking at on the screen. The entire study was managed using LimeSurvey, an open-source online survey tool, hosted on our secure servers. The websites were also hosted on our laboratory's secure servers, and were only accessible from our local area network.

5.1.3 Materials and Design Two versions of the website were used: H1 (HVA with a DV which is less secure) and L2 (LVA but with an OV which is more secure). Ten information retrieval tasks (five per website) were given to participants,

Table 2. Likert Scale Security Questions

*	This is a legitimate website offering genuine services.
*	This is a fake website intended to trick users into using it.
*	I am confident enough in the authenticity of the website to create an account.
*	I am doubtful about the authenticity of the website to create an account.
	I would feel safe to purchase something from this website.
	I would feel it was dangerous to purchase something from this website.
*	I would feel it was safe to store private information on this website.
*	I would feel it was dangerous to store private information on this website.
	This website is secure from hackers.
	This website has been taken over by hackers.
*	The website is trustworthy.
*	The website is suspicious.
	The website would guard your data carefully.
	This website might allow unauthorized access to your data.

* Used in Study 2

in random order. Examples of these tasks are: "How many beaches are located in the Gold Coast," "Does the Gold Coast require pet dogs to be registered," and "What breakfast restaurant is highly recommended and won an award?" None of the information retrieval questions targeted website or personal security.

An informed consent and demographic questionnaire were administered to determine the participants' background information (e.g. age, gender, and education). The System Usability Scale (SUS; [4]) and the Visual Aesthetics of Websites Inventory—Short version [VisAWI-S; 29] scales were used for perceived measures of usability and visual appeal, before and after website use, respectively. The objective usability in the form of per task performance measures were: Task completion time in seconds with a maximum three minutes and success (pass if the answer is correct and within time limit). Both the SUS and VisAWI-S used the 5-point Likert format, where a score of 1 means 'strongly disagree' and 5 means 'strongly agree'.

We created our own questionnaire to evaluate perceived security (Table 2). There were 14 items: seven positively-phrased questions about website security (e.g. "I would feel safe to store private information on this website") and seven negatively-phrased variants (e.g. "I would feel it was dangerous to store private information on this website"). This was to help ensure that participants were attentive and deliberate in their responses. Like the SUS and the VisAWI-S, each question used a 5-point Likert scale.

This study adopted a within-group design, where all participants saw both websites and filled in each questionnaire. The interaction with the websites was counterbalanced, to avoid order effects. Participants were only told to evaluate the websites with their honest opinions, with no mention of security, in an effort to avoid priming, biasing, or otherwise influencing participant experience and opinions.

5.1.4 Procedure Each participant did the experiment separately in one-on-one sessions with the researcher. Each session started with a briefing on the study's purpose, and an explanation of how the eye-tracker works. Next, participants signed informed consent forms before beginning the study. The study started with the demographic questions before prompting the participants to click a link which would open one of the two websites in a different tab. At this stage, participants had 30 seconds to browse through it before being prompted to return to the survey and fill in the visual appeal, usability, and security scales. Then participants were asked to go back to the website they had just interacted

with and complete five information retrieval tasks, one at a time. Participants completed all five tasks before proceeding with the same visual appeal, usability, and security questionnaires about their experience, after having used the website for up to 15 minutes. This sequence of browsing, rating, doing five new information retrieval tasks, and rating again was repeated for the other website but for a new set of five questions. Participants then completed the last set of post-task questionnaires (VisAWI-S, SUS, and Table 2) about their experience and which website they preferred for visual appeal, usability, and security. The researcher asked them summary questions that reiterated the preference questionnaire and asked for additional comments or feedback.

5.2 Results

First, we examine the eye-tracker results, to gain an understanding of what participants were looking at during their interactions with the websites. Second, the results from the VisAWI-S, SUS, and security scale were plotted, using beanplots to gain a general understanding of the data. Third, we checked the normality and homogeneity of variance assumptions. Thus, we proceeded to use parametric statistics reported in the Statistical Analysis section below.

5.2.1 Eye-Tracker During the pre-use browsing task, none of the participants fixated on, or interacted with the browser indicator next to the URL, for both websites. Instead, for both websites, users spent their time examining the body of text in the center of the page, the small print at the bottom of the page to see that the copyright year was current, the title and logo at the top left, and the website's menu bar at the top, often pausing to examine the menu options in the dropdown that appeared when participants hovered over an option. For the information retrieval tasks, participants focused on searching for the answers within the websites. None of them fixated their gaze on the browser indicators or the URL during these tasks. *This suggests that the certificate was not a factor that influenced user perceptions of security.* Therefore, the only other factor that could have caused the difference in results is visual appeal.

5.2.2 Beanplots Data were graphed using beanplots [18] to gain an understanding of any emerging trends. Beanplots are similar to box plots, but they also show the distributions on both sides of the vertical bar. Beanpots visually present the population spread which allows for more accurate conclusions to be drawn. Beanplots were created to gain a general understanding of the data, with pre- and post-use for both websites being depicted in each beanplot. Fig. 4a shows the results for visual appeal ratings, Fig. 4b shows the security results, and Fig. 4c shows the usability results. In each figure, the grey beans are the pre-use ratings, and the white are post-use. The first columns on the left represent the Hva website, and the ones on the right represent the Lva website. The data appeared normally distributed.

Visual appeal and usability both show the same trend: Pre-use ratings (the grey beans) are a little higher than post-use ratings (the white beans). This seems to suggest that participants had somewhat lower perceptions of usability and visual appeal after use. The statistical results section, below, examines the significance of this difference. The opposite trend occurred for security, where post-use ratings seem to be higher than the pre-use ratings, especially in the Lva website. However, the security rating of the Hva website is virtually identical pre-and post-use in Fig. 4b, suggesting that this may not be a significant difference. Given that the certificate indicators did not change throughout this study, the fluctuations in security ratings may be attributable to other factors, as suggested by our hypotheses. This topic is further discussed in the Statistical Analysis section below.

When examining the differences between the means of visual appeal, security, and usability, as seen in Table 3, we see that visual appeal, security, and usability appear to be closely rated, across all websites, pre- and post-use. *Specifically, we note that even though the Lva website is regarded as more secure by offering more identity information in the certificate, it is being rated as less secure than the Hva website, which has a DV certificate.* This trend continues after use as well,

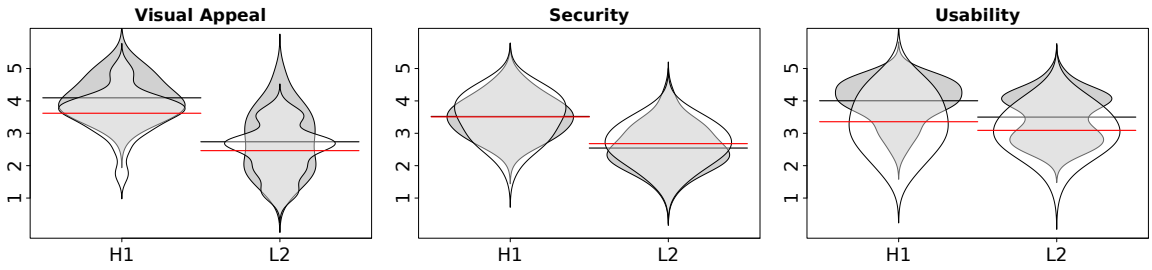


Fig. 4. Likert scale ratings (1 is 'strongly disagree'; 5 is 'strongly agree') before and after use for visual appeal, security, and usability questionnaires: HVA/DV (H1) and LVA/OV (L2) websites. The dark-coloured beans represent *pre-use* ratings, and the light-coloured beans *post-use*. The horizontal lines indicate each bean's mean: black for pre-use, and red for post-use.

Table 3. Means ($\bar{X}$) and standard deviations (SD) of Likert scale responses (1 = 'strongly disagree'; '5 = strongly agree') to the visual appeal, security, usability questionnaires, before and after use: HVA/DV (H1) and LVA/OV (L2) website

	Pre-Use				Post-Use			
	H1		L2		H1		L2	
Questionnaire	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Visual Appeal	4.10	0.54	2.74	0.97	3.62	0.76	2.46	0.78
Security	3.52	0.57	2.54	0.66	3.51	0.70	2.68	0.60
Usability	4.00	0.64	3.50	0.73	3.36	0.88	3.09	0.67

which supports the hypothesis that the HVA website would be rated as more secure, regardless of the actual security level. The significance of these differences is tested in the next section.

5.2.3 Statistical Analysis We checked the normality and homogeneity of variance assumptions, to ensure that we used the proper statistical tests (i.e. parametric or non-parametric). Normality was tested using Shapiro-Wilk [35, 40], and Levene's test [31] was used to examine the homogeneity of variance assumption. None were violated. Thus, we proceeded to use parametric statistics.

For t-tests with significant results we report Cohen's d effect sizes. We adopt Cohen's benchmarks for effect sizes: a d value of 0.2 is interpreted as a "small" effect size, 0.5 is "medium", and 0.8 is "large" [7]. Cohen's d is the difference in means divided by the pooled standard deviation, the latter of which is calculated here as:

$$\sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

Verifying the differences between the two websites. To verify that visual appeal was significantly better in the HVA website, we ran paired, one-sided t-tests. We ran one-sided tests since we expected the HVA conditions to be rated as more secure. As expected, visual appeal was significantly higher in the HVA website pre- ($t(20) = 6.27, p < .001, d = 1.72$) and post-use ($t(20) = 5.98, p < .001, d = 1.55$) than in the LVA website. Both pre-use and post-use d values (1.72, 1.55) exceed the threshold for a "large" effect size. The objective usability of the two websites was not significantly different,

Table 4. Pearson's correlation coefficients (r) between pre- and post-use ratings of perceived visual appeal, security, and usability, for the Hva and Lva websites

	Pre-Use vs. Post-Use	
Questionnaire	Hva	Lva
Visual Appeal	.75 ***	.77 ***
Security	.80 ***	.77 ***
Usability	.57 *	.39 ns

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

as the differences in average time per task ($t(20) = 0.78$, $p = .44$) and the average number of tasks completed correctly ($t(20) = 0.70$, $p = .49$) were both insignificant. This suggests that participants used the websites with similar efficacy and any differences in perceived usability may be due to non-usability factors. The average time per task and the average number of successfully completed tasks were negatively correlated in both the Hva ($r = -.64$, $p = .002$) and Lva ($r = -.62$, $p = .004$) websites. This suggests that participants who had difficulty on a task had to work on it for longer before moving on to the next task.

Experimental factors. To test the impact of visual appeal on security, we ran two one-sided, paired t-tests, where the expectation was that the Hva website would be rated as having higher security ratings (regardless of actual security). As expected, security was rated higher on the Hva website both pre- ($t(20) = 6.08, p < .001, d = 1.58$) and post-use ($t(20) = 4.11, p < .001, d = 1.26$). Both pre-use and post-use d values (1.58, 1.26) exceed the threshold for a "large" effect.

To examine the impact of use and of interacting with the website, we ran two paired, two-sided t-tests on perceived security ratings pre- and post-use. Use did not alter people's opinions of security of the Hva ($t(20) = 0.15, p = .88$) or the Lva ($t(20) = -1.43, p = .17$) website.

To verify that visual appeal also affected usability, as per the second hypothesis, we ran two one-sided, paired t-tests. Perceived usability between the two websites differed significantly before use ($t(20) = 3.13, p = .003, d = 0.74$) but not after ($t(20) = 1.69, p = .05$). The d pre-use d value exceeds the threshold for a "medium" effect size.

5.2.4 Implications for the Hypotheses Therefore, the results of Study 1 provides evidence for the first hypothesis, that what is beautiful is secure. The appealing website yielded higher levels of perceived security than the visually unappealing website, even though the less appealing one had verified identity ownership and the correct URL. Based on these results, we could conclude that visual appeal drove user perceptions of security. Furthermore, the results of Study 1 also provide evidence for the second hypothesis, that visual appeal was relied on for the judgments of website usability, where the website that was higher in visual appeal also scored higher in usability, particularly pre-use.

5.2.5 Correlations For purposes of comparison to other correlational studies in the literature (e.g. [53]), and to verify the relationship of visual appeal, security, and usability, we ran a series of Pearson correlations, the results of which can be seen in Tables 4 and 5, and visualized in Fig. 5. First we compare pre-use and post-use ratings for the three questionnaires by website (Table 4). As expected, pre-use visual appeal ratings for both the Hva and Lva websites were strongly and positively correlated with their post-use counterparts. The same was true for pre- and post-use ratings of security. For usability, strong positive correlations were observed only with Hva websites.

Table 5. Pearson's correlation coefficients (r) between perceived visual appeal and security, pre- and post-use, for the HVA and LVA websites

			Visual Appeal	
			Pre	Post
Security	HVA	Pre	.72 ***	.39 ns
		Post	.62 **	.37 ns
	LVA	Pre	.63 **	.32 ns
		Post	.45 *	.45 ns

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

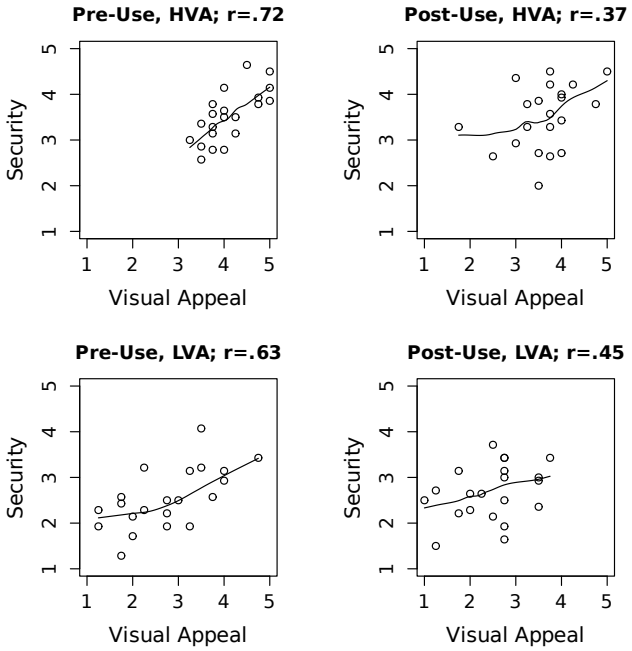


Fig. 5. Pre-pre and post-post Pearson's correlation coefficients (r) from Table 5: 1 = 'strongly disagree; 5 = 'strongly agree'

Table 5 shows correlations between visual appeal and security, pre- and post-use, for both websites: HVA and LVA. For the HVA website, pre-use visual appeal was strongly and positively correlated with pre-use and post-use security. Similarly, for the LVA website, pre-use visual appeal was strongly and positively correlated with pre-use and post-use security.

Visual appeal and usability for the HVA website were also strongly and positively correlated pre-use ($r = .51$, $p = .019$) and post-use ($r = .53$, $p = .014$). Similarly, for the LVA website, visual appeal was significantly and positively correlated to usability both pre- ($r = .44$, $p = .04$) and post-use ($r = .47$, $p = .03$). These results support the beanplot and statistical analysis results from the previous two sections.

5.2.6 Choice Results Each participant was asked to identity which website they would consider: the most visually appealing, the most usable, the most secure, which they would purchase from, which they would log in to, and which they thought was the 'real' one. *100% of the participants chose the HVA website for all of these questions.* Many participants mentioned that the HVA website was more "trustworthy". The LVA website was described as "weird", where the colours and pictures felt "off". Participants said that the LVA website had a poor, bad-looking design. Only one participant said that the websites appeared to have the same security levels, yet they said they would still choose the HVA one over the LVA one for any task that they had to accomplish. These results also support the hypothesis that the HVA website yields higher ratings of security, regardless of the actual website security levels.

5.3 Discussion

Taking all the results together, from the eye-tracker, the beanplots, the statistical and correlational analysis, there is strong evidence that supports the hypothesis that visual appeal largely influences perceived security. It also provides evidence that supports previous findings in the literature (e.g. [54]), that visual appeal influenced perceived usability.

More importantly, security was rated significantly higher on the HVA website, before and after use. This shows us that visual appeal did strongly influence participants' opinions of security. Therefore, *what is beautiful is secure* holds. These findings were further supported by the correlation analysis, which are also in alignment with previous work in the field (e.g. [53]).

Every participant stated that the HVA website was indeed higher in visual appeal, more usable, and more secure. This is dangerous, as it demonstrates a willingness on behalf of users to hand sensitive information over to a website as long as it has high visual appeal, regardless of its actual level of security.

5.4 Limitations

While the results of Study 1 provide evidence to support the hypotheses, it does have some limitations. The study demographic was homogeneous, with most participants being young and educated. It also had a relatively small number of participants (only 20). Moreover, Study 1 focused on two web certificates, whereas there are three website certificate types, with the absence of a certificate accounting for a fourth certificate condition. Therefore, there was a need to do a more detailed analysis and follow-up study that would examine more website conditions along with polling a larger and more diverse participant group. Another potential limitation is that the security scale we created and used was not validated beforehand. The scale may have also been too long, as some participants found some questions redundant. For Study 2, the scale was shortened (please see Section 6.1.2 (Materials and Design) in Study 2 for more details).

5.5 Conclusions

The results of Study 1 indicate that visually appealing websites are rated as more secure. We were able to arrive at this conclusion by keeping as many factors constant in Study 1. Specifically, there were two opposing visual appeal levels (high and low), and two seemingly identical security levels (OV and DV), since these are indistinguishable without interaction with the indicators. Therefore, the difference in ratings would most readily be attributed to visual appeal. On the basis of the literature, we suspected people would not interact with indicators, and our eye-tracking results confirm this. Without interaction, the security indicators appeared identical, meaning that visual appeal was the only factor that affected user decisions. In broader usage of the web, however, users may encounter visually distinct security indicators. Hence, to get a better understanding of the influence of visual appeal on security, Study 2 included all four security conditions and both visual appeal variants. The details of Study 2 are in the next section.

6 Study 2: Online Study

The results of Study 1 confirmed the hypothesis that visual appeal was influencing security decisions, and also the hypothesis that it influenced usability. However, that study had some limitations: there were only 20 participants, and we used only two certificate levels (OV and DV). The aim of Study 2 was to examine more cases in order to enable a more detailed analysis of the effect of visual appeal on security and usability. To see if our hypothesis held with a larger and potentially more diverse sample of participants, and to separately examine the effects of all four certificate levels (an exhaustive list), we devised a second, online study, which recruited participants using crowdsourcing recruitment platform MTurk through CloudResearch [25]. The study was cleared by our research ethics board.

In Study 2, we focused on further testing both hypotheses. As previously mentioned, our primary hypothesis was that a highly appealing website would yield higher levels of perceived security than a visually unappealing website, regardless of what security indicator is shown. If people rated the more visually appealing website as more secure, regardless of its actual security level, then we could conclude that visual appeal drove user perceptions of security. Our secondary hypothesis was that participants will rate the more visually appealing websites as more usable, even though the websites' actual usability levels were identical.

In Study 2, all four certificate security levels were used in combination with the two visual appeal levels (HVA and LVA) to give a total of eight conditions: H3, H2, H1, H0, L3, L2, L1, L0. As mentioned earlier, EV is the certificate offering the most assurances to users, but OV is more widespread, and we were interested in examining whether users distinguish between it and the less secure DV. In Study 2, we use the legacy EV certificate indicator (seen in Fig. 3a) and the current indicators: OV, DV, and 'no certificate' (Figs. 3b–3d). We use the legacy EV indicator for purposes of comparison with current indicators.

A main threat to website security is fraudulent (e.g. phishing) websites that have similar appearance and content to real websites, and they have URLs chosen to mislead users. They can only be distinguished from genuine sites by the certificates or by knowing and carefully checking the URLs. The legacy EV and current OV (security levels 3 and 2, respectively) website variants were given the website's real URL, and the DV and no certificate (levels 1 and 0, respectively) variants were given a fake but plausible URL, to mimic real-life situations and phishing attacks. The URL/certificate indicator combinations shown to users can be seen in Fig. 3 (see Section 3). Based on the results of Study 1, where participants did not look at or interact with certificate interfaces, and since we wanted the same participant to rate 4 conditions, participants in Study 2 examined screenshots rather than interactive websites.

6.1 Method

6.1.1 Participants 186 people (97 males, 89 females) volunteered for this study. Out of these, 53 were aged 20–29, 81 were aged 30–39, 35 were aged 40–49, and 17 were aged 50 or over. 53 had a high school diploma, 108 had an undergraduate degree, 22 had a master's degree, 3 had a PhD or professional school (medical, dental, legal, etc.) qualification. 40 had backgrounds in computer science or engineering; 29 in sociology, psychology, or cognitive science; 24 in biology, chemistry, or medicine; 18 in history, art, education, politics, or linguistics; 18 in business; 11 in mathematics, statistics, or physics; 2 in law, criminology, or forensics; 1 in construction; the rest had no academic background or did not wish to disclose this information. Participants individually completed the survey online, with each session lasting approximately 15 minutes. All participants were English-speakers.

6.1.2 Materials and Design The questionnaire was hosted on our laboratory's secure server using LimeSurvey, and distributed by CloudResearch. To better ensure careful responses, we accepted only those participants with a 99%

approval rating and above, and more than 1000 Human Intelligence Tasks (HITs) approved. We also included measures in the survey to filter out bots, or those who were not paying attention. Inspiration for such measures came from work done by [5]. We provided a short math question ("What is 2 + 2?") in the form of an image, so that it could not easily be answered via a script. We also added several text boxes with no accompanying instructions; several participants entered comments in these text boxes that were either out-of-context, random-seeming, or that otherwise made no sense. We also restricted it to participants in the USA. The SUS and secsurity scales have negative questions to help identify those who are not carefully reading the questions. A total of 14 participants were removed as a result of our screening process.

An informed consent was administered first. Then, a demographic questionnaire was administered to determine the participants' background information (e.g. age, gender, and education). The same scales from Study 1 were administered for visual appeal and usability. The security questionnaire was shortened to eight questions (four positive; four negative) to save time. This was done since brief analysis showed that it was sufficient (no significant differences between the long and short results). Each question in all of these scales used the Likert format (5-point system, from 'strongly disagree' to 'strongly agree'). The final post-task preference questionnaire involved 7 choice questions where participants had to select one of the four websites that they thought was the *most* visually appealing, usable, secure, which one they would purchase from, log in to, recommend, and which they thought was real.

To avoid fatigue and confusion during the study, the eight conditions were split into two groups in Study 2. Each group had one of each certificate, shown in Fig. 3, and both levels of visual appeal. In the first group, the Lva websites had the two highest security levels (EV and OV) and the two Hva websites had the lowest security levels (DV and no certificate): L3, L2, H1, H0. We call this group Mismatched henceforth, as *high* visual appeal is paired with *low* security, and vice versa. In the second group, the Hva websites had the highest security levels, and the Lva websites had the lowest security levels (DV and no certificate): H3, H2, L1, and L0. This group we call Matched.

Half of the participants were put into the first group, and half were assigned to the second. Each person saw each certificate level once, but saw two instances of the Hva and two instances of the Lva website. This was done to eliminate any confusion about certificate levels and any learning effects therein. The exposure to the websites within each group was randomized, to avoid order effects. Participants were only told to evaluate the websites with their honest opinions, with no mention of security, in an effort to avoid priming, biasing, or otherwise influencing participant experience and opinions.

6.1.3 Procedure Each participant did the experiment online, at their own convenience. After the participants answered demographic questions, they were shown four websites in randomized order, according to the groups mentioned in the previous section. For each of the four websites, participants had a screenshot of the homepage to examine and were asked to fill in the visual appeal, usability, and security scales for the screenshot. Participants then completed the final post-task 7-item preference questionnaire.

6.1.4 Data Analysis Data were graphed using beanplots to gain an understanding of any emerging trends. We checked the normality and homogeneity of variance assumptions, to ensure that we used the proper statistical tests. We then proceeded to examine the data statistically to test the hypotheses.

6.2 Results

6.2.1 Beanplots

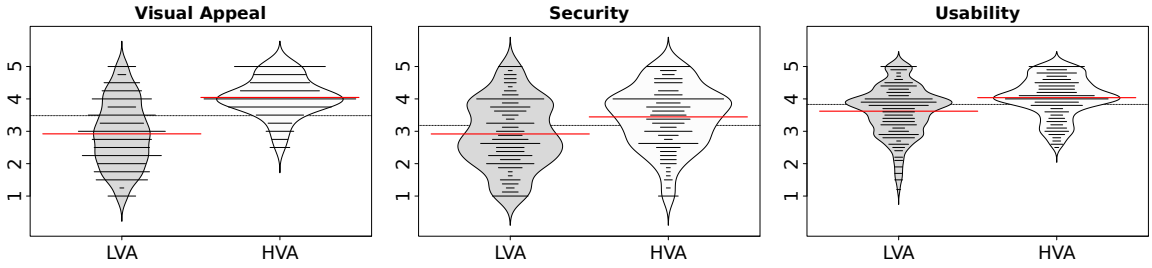


Fig. 6. Mismatched ratings (1 = 'strongly disagree'; 5 = 'strongly agree') aggregated by visual appeal level. The red lines show group means, and the dotted lines stretching across the plot show overall means.

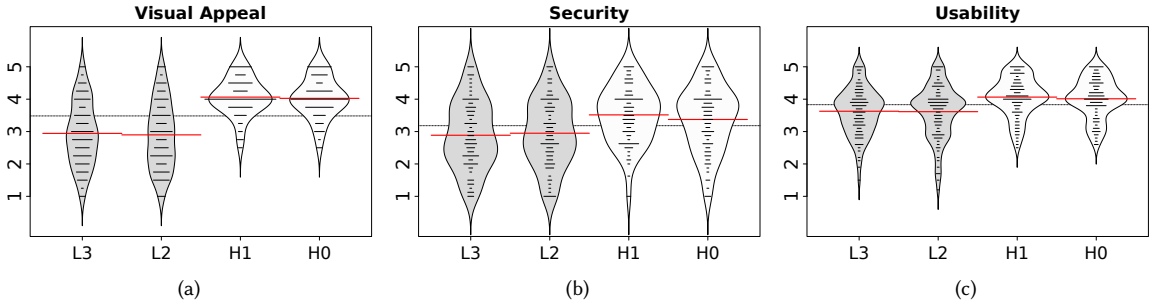


Fig. 7. Mismatched ratings (1 = 'strongly disagree'; 5 = 'strongly agree') for each website condition. The red lines show group means, and the dotted lines stretching across the plot show overall means.

Table 6. Mean ratings per website, Mismatched and Matched groups

	Mismatched				Matched			
	L3	L2	H1	H0	H3	H2	L1	L0
Visual Appeal	2.94	2.90	4.06	4.03	4.04	3.92	2.73	2.73
Security	2.89	2.95	3.52	3.37	3.64	3.51	2.71	2.41
Usability	3.63	3.61	4.06	4.01	3.91	3.94	3.48	3.42

Mismatched: Beanplots of participant ratings of visual appeal, security, and usability were created to gain a general understanding of the data. In Fig. 6, responses from the four conditions (L3, L2, H1, and H0) were aggregated by visual appeal to help highlight the effect of the websites' respective visual appeals: in each plot, the first bean represents the two Lva websites, and the second bean represents the two Hva websites. Then, Figs. 7a–7c separate the conditions to look for effects of certificate security level. The means for Mismatched can be seen in Table 6.

As seen in Fig. 6, *the Lva website conditions were rated as less visually appealing, less secure, and less usable than the Hva*. This is indicated by the difference between the beans' respective means bars, and also by the shape of the beans, where the Hva beans tend to be slightly heavier on top.

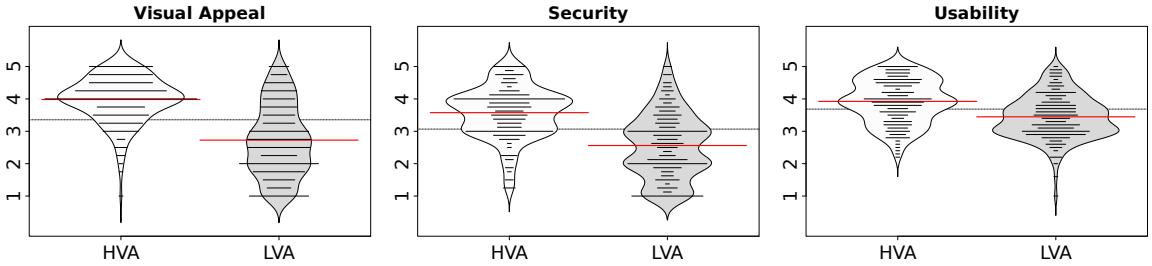


Fig. 8. Matched ratings aggregated by visual appeal level

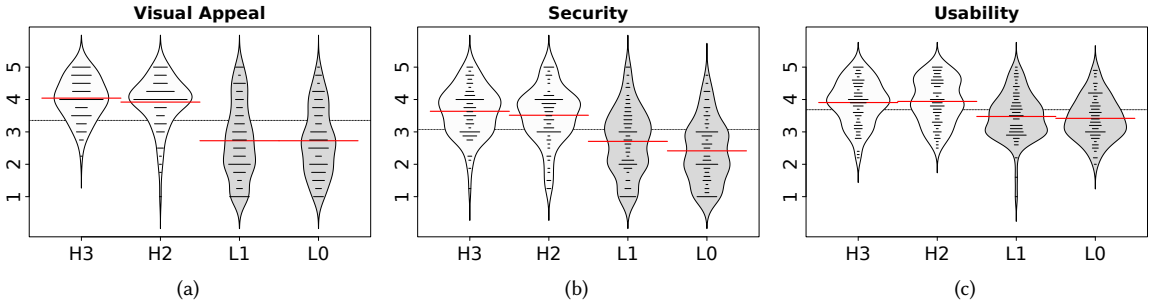


Fig. 9. Matched ratings for each website condition

When viewing the conditions separately (Fig. 7), it is apparent that the pattern persists across all certificate levels. Specifically, the means for the Hva websites are higher across all three factors: visual appeal, security, and usability. Further inspection shows that the means for the two Hva websites and the two Lva websites were nearly identical for visual appeal (Fig. 7a) and usability (Fig. 7c). However, in addition to the difference between the Lva and Hva websites, there was a small difference between H1 and H0 in the security ratings (Fig. 7b), where H1 was rated as slightly more secure than H0.

Matched: Fig. 8 shows beanplots for Matched aggregated by visual appeal, and Fig. 9 depicts each condition separately. The means are shown in Table 6.

Fig. 8 shows that the Lva website conditions were rated as less visually appealing than the Hva, less secure, and less usable. Again, the mean bars for Hva are higher than Lva, and the Hva beans are also heavier on top.

When examining the Matched beans separately for each condition (Fig. 9), the pattern persists across all certificate levels. Specifically, the means for the Hva websites are higher across all three factors: visual appeal, security, and usability. Further inspection shows that the means for the two Hva websites and the two Lva websites were identical or nearly identical for visual appeal (Fig. 9a), between H3 and H2 for security (Fig. 9b), and usability (Fig. 9c). There was a small difference between L1 and L0 in the security ratings (Fig. 9b), where L1 was rated as slightly more secure than L0.

It seems that there is a pattern across both groups that supports both hypotheses: Users perceive more visually appealing websites as more secure and more usable, *even when the visually appealing website's security indicator suggests that the website is not secure.* In the next section, the statistical significance of these results is examined.

Table 7. Tukey HSD multiple comparisons of **security means** with 95% family-wise confidence: Mismatched and Matched. The "Diff." column shows the difference in security means between the two websites.

Mismatched				Matched			
Conditions	Diff.	p	d	Conditions	Diff.	p	d
L2 − L3	0.06	ns	-	H2 − H3	−0.12	ns	-
H1 − L3	0.63	***	0.65	L1 − H3	−0.93	***	−1.05
H0 − L3	0.49	**	0.48	L0 − H3	−1.22	***	−1.42
H1 − L2	0.57	***	0.59	L1 − H2	−0.81	***	−0.87
H0 − L2	0.42	*	0.42	L0 − H2	−1.10	***	−1.21
H0 − H1	−0.14	ns	-	L0 − L1	−0.29	ns	-

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

Table 8. Tukey HSD multiple comparisons of **usability means** with 95% family-wise confidence: Mismatched and Matched. The "Diff." column shows the difference in usability means between the two websites.

Mismatched				Matched			
Conditions	Diff.	p	d	Conditions	Diff.	p	d
L2 − L3	−0.01	ns	-	H2 − H3	0.03	ns	-
H1 − L3	0.44	***	0.66	L1 − H3	−0.43	***	−0.63
H0 − L3	0.38	***	0.57	L0 − H3	−0.49	***	−0.74
H1 − L2	0.45	***	0.63	L1 − H2	−0.46	***	−0.67
H0 − L2	0.40	***	0.55	L0 − H2	−0.52	***	−0.79
H0 − H1	−0.05	ns	-	L0 − L1	−0.06	ns	-

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

6.2.2 Statistical Analysis We checked the normality and homogeneity of variance assumptions, and found that none were violated. Thus, we proceeded to use parametric tests to test our hypothesis: ANOVAs with follow-up Tukey HSD tests, and Pearson's correlation coefficients. Pearson's correlations were examined for purposes of comparison with prior studies in the literature on visual appeal and usability that were correlational in nature.

For ANOVA tests with significant results we provide η_p^2 effect sizes. For Tukey HSD tests with significant results we present Cohen's d effect sizes (see Tables 7 & 8).

Mismatched: Our first hypothesis was that a highly appealing website would yield higher levels of perceived security than a visually unappealing website, regardless of what security indicator is shown. If people rated the more visually appealing website as more secure, regardless of its actual security level, then we could conclude that visual appeal drove user perceptions of security. In the case that there are two Hva websites and two Lva, we predict that the Hva websites will be rated as more secure, even if their security levels are worse.

A one-way, four-group ANOVA (within-subjects) was computed. Significant main effects ($F(3, 388) = 9.59$, $p < .001$, $\eta_p^2 = 0.07$) were found for security scores between the L3, L2, H1, and H0 conditions. Tukey HSD multiple comparisons of means with 95% family-wise confidence were calculated (see Table 7). In Tables 7 and 8, every comparison between high and low visual appeal was significantly different ($p < .001$ in most cases), with Hva always being rated as more secure than Lva. In Mismatched, the two Hva conditions were rated as the most secure despite them being the least

secure, providing supporting evidence for the hypothesis that visual appeal is guiding security judgments, regardless of actual security levels. The only two non-significant differences were (1) between the Lva with OV and Lva with EV (L2 & L3), and between Hva with no certificate and Hva with DV (H0 & H1). When participants were presented with websites that look the same but with varying degrees of security, there is no evidence that they rated them differently in terms of security.

Our second hypothesis was that participants will rate the more visually appealing websites as more usable, even though all four websites' actual usability levels are identical. Significant main effects were found ($F(3, 388) = 11.90$, $p < .001$, $\eta_p^2 = 0.08$) for usability scores between the four websites in the Mismatched condition. Tukey HSD results (Table 8) show that the Hva websites were always rated as more usable than the Lva. The usability ratings of the two Hva, and the two Lva also were not significantly different from each other. This evidence strongly supports our second hypothesis that visual appeal influences usability ratings.

Matched: In seeking an answer to our first hypothesis, a one-way four-group ANOVA (within-subjects) was computed for the security ratings of the Matched group. Significant main effects were found between the four websites ($F(3, 348) = 39.49$, $p < .001$, $\eta_p^2 = 0.25$). Tukey multiple comparisons of means with 95% family-wise confidence were then calculated. Once again, as seen in Table 7, every pair that compared high with low visual appeal categories was significantly different ($p < .001$ in each case), with Hva always being rated as more secure than Lva. The only two non-significant differences were (1) between the Hva with OV and Hva with EV (H2 & H3), and (2) between Lva with no certificate and Lva with DV (L0 & L1)—cases where there were actual differences in security, but no difference in visual appeal. This is consistent with our first hypothesis.

For the second hypothesis, we predicted that usability would be rated higher when visual appeal was higher and that usability would be rated lower when visual appeal was lower. Significant main effects were found ($F(3, 348) = 14.79$, $p < .001$, $\eta_p^2 = 0.11$) for usability scores between the four websites in the Matched condition. Once again, the Tukey HSD results (in see Table 8) show that the Hva websites were always rated as more usable than the Lva. Similar to the Mismatched group, the usability ratings of the two Hva, and the two Lva also were not significantly different from each other. This provides further evidence in support of our second hypothesis that visual appeal seems to be dictating the usability ratings.

Comparing across Mismatched and Matched: A series of pairwise between-group t-tests were conducted to compare security means *across* Mismatched and Matched. The tests are one-sided where the visual appeal levels differ, where the Hva security means are predicted to be greater than the Lva means; the tests are two-sided otherwise. The Bonferonni-corrected results are shown in Table 9. In all cases where visual appeal levels differ between the two websites, the Hva website is rated as significantly more secure ($p < .001$) than the Lva website. This is true *even in cases where the Hva security level is lower*: L3-H2 ($t(182.5) = -4.51, p < .001$) and H0-L1 ($t(181.75) = 4.53, p < .001$). Of these, H0-L1 is the most remarkable: H0 is rated as more secure despite its security indicator explicitly reading "Not secure", and the L1 indicator showing a lock symbol.

6.2.3 Correlations To further explore the relationship of visual appeal on security and usability, Pearson's correlations were computed; the results are provided in Table 10. In all the correlations explored, there was a reasonably high ($r = 0.3$–0.65), positive, and statistically significant ($p < .01$ in one case; $p < .001$ in the rest) correlation. These suggest that visual appeal is related to both security and usability ratings, pre-use.

Table 9. Pairwise between-group t-tests of **security means** with Bonferonni correction, comparing Mismatched against Matched. The "Diff." column shows the difference in means (Mismatched−Matched) between the two websites.

		Mismatched											
		L3			L2			H1			H0		
		Diff.	p	d	Diff.	p	d	Diff.	p	d	Diff.	p	d
Matched	H3	−0.75	***	−0.83	−0.69	***	−0.76	−0.12	ns	-	−0.26	ns	-
	H2	−0.63	***	−0.66	−0.57	***	−0.59	0.00	ns	-	−0.14	ns	-
	L1	0.18	ns	-	0.24	ns	-	0.81	***	0.86	0.67	***	0.67
	L0	0.47	*	0.47	0.53	**	0.53	1.10	***	1.20	0.96	***	0.98

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

Table 10. Per-condition Pearson's correlation coefficients (r) between (i) visual appeal (VA) and security, and (ii) VA and usability: Mismatched and Matched groups

Mismatched			Matched		
	VA & Security	VA & Usability		VA & Security	VA & Usability
L3	.49 ***	.43 ***	H3	.59 ***	.58 ***
L2	.62 ***	.60 ***	H2	.43 ***	.58 ***
H1	.35 ***	.56 ***	L1	.62 ***	.54 ***
H0	.30 **	.65 ***	L0	.49 ***	.43 ***

**: $p < .01$ ***: $p < .001$

6.2.4 Choice Questions Of the four websites they saw, participants were asked to select which best matched each of the following criteria:

(1) the most visually appealing,
(2) the most secure,
(3) the most usable,
(4) the one they would purchase from,
(5) the one they would log in to,
(6) the one they would recommend, and
(7) the one that was a 'real' website.

A chi-square test was calculated first, followed by family adjusted pairwise comparisons for each question (Tables 11 and 12).

Mismatched: For each of the seven questions, H1 was the most popular response in Mismatched (Table 11), followed by H0, and then L3 and L2 which were not significantly different. This means that the Hva website with a DV was rated as the most visually appealing, more secure, and most usable. It was also chosen the most when participants were asked which website they would make a purchase from, which one they would log in to, which one they would recommend to others, and which one they thought was the authentic website. This means that most participants chose

Table 11. Choice questions, MISMATCHED: Choice frequencies with chi-square test, and family-adjusted pairwise comparisons

	Choice frequencies					Family-adjusted pairwise comparisons					
Question	L3	L2	H1	H0	χ^2	L3-L2	L3-H1	L3-H0	L2-H1	L2-H0	H1-H0
Visually appealing	5	2	66	25	106.49 ***	ns	***	***	***	***	***
Secure	9	4	67	18	102.41 ***	ns	***	ns	***	**	***
Usable	17	6	52	23	47.22 ***	*	***	ns	***	**	**
Purchase from	16	8	50	24	40.61 ***	ns	***	ns	***	**	**
Log in to	17	7	56	18	57.02 ***	*	***	ns	***	*	***
Recommend	16	8	50	24	40.61 ***	ns	***	ns	***	**	**
Real	15	7	55	21	54.65 ***	ns	***	ns	***	*	***

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

Table 12. Choice questions, MATCHED: Choice frequencies with chi-square test, and family-adjusted pairwise comparisons

	Choice frequencies					Family-adjusted pairwise comparisons					
Question	H3	H2	L1	L0	χ^2	H3-H2	H3-L1	H3-L0	H2-L1	H2-L0	L1-L0
Visually appealing	59	22	3	4	93.36 ***	***	***	***	***	***	ns
Secure	56	27	3	2	88.27 ***	**	***	***	***	***	ns
Usable	57	22	7	2	84.09 ***	***	***	***	**	***	ns
Purchase from	55	25	8	0	38.61 ***	**	***	***	**	***	**
Log in to	56	24	5	3	82.27 ***	***	***	***	***	***	ns
Recommend	56	24	4	4	82.18 ***	***	***	***	***	***	ns
Real	55	26	4	3	81.36 ***	**	***	***	***	***	ns

*: $p < .05$ **: $p < .01$ ***: $p < .001$ ns: $p > .05$

the most secure HVA website—even when there were more secure LVA options. Participants even chose the image that explicitly said "Not secure" as being more secure than ones with valid certificates.

MATCHED: H3 was the most popular response in MATCHED (Table 12), followed by H2, and L1, and L0. In other words, most participants in MATCHED thought that the HVA website with an EV certificate was the most visually appealing and most secure. This website was also chosen as the one they would make a purchase from, the one they would log in to, the one they would recommend to others, and the one they thought was the most authentic. However, the usability of H3 was not statistically different from H2. Therefore, in MATCHED, most participants chose the visually appealing website with the highest security—in this case, it actually did have the highest security certificate. In fact, the most popular case was the legacy EV case, and despite it being discontinued, it was interpreted as representing higher security.

6.3 Discussion

Altogether, the findings of Study 2 support the first hypothesis, *What is beautiful is secure.* Across the statistical and correlational analysis of website ratings, and also in the choice questions, participants consistently rated visually appealing website as more secure than their visually unappealing counterparts, *even when the visually appealing website featured no certificate, and was explicitly labelled as "Not secure".* We think this latter finding best demonstrates the

power of visual appeal and its attendant effects on feelings of security: a strong signal of a website's actual insecurity (the "Not secure" label) was overshadowed by judgments of aesthetic qualities. Further discussion on the certificate indicators and the halo effect are presented in the General Discussion section, below.

Additionally, in support of our second hypothesis, participants rated visually appealing websites are more usable as well.

6.4 Limitations

One limitation of our design in Study 2 is that participants were not given the opportunity to use the websites or interact with the web certificates. However, as mentioned earlier, previous findings suggest that they would have done so, even if the opportunity was presented to them. Future research may seek to examine these hypotheses on familiar websites and on mobile devices to mimic real life scenarios more readily.

6.5 Conclusions

The findings of Study 2 outline that objective security may play a role in people's opinions of website security, even in the absence of interaction with the browser indicators. However, the impact of visual appeal far surpasses the impact of the security indicators, and arguably serves as a primary filter for security. Taken together with the results of Study 1, we believe that we have strong evidence that *What is beautiful is secure.* These findings supplement previous findings that visual appeal affects usabilty (e.g. [54]) and trust (e.g. [34]). Further investigation should occur to examine what other factors are impacted by visual appeal.

7 General Discussion

The most reliable indication of a website's identity—and its connection to a trusted entity—comes from web certificates. The assurances they provide cannot be perfect, but nonetheless they remain worth having. Web browsers provide a summary of this information in a lock icon placed beside the URL bar. Clicking on the icon brings up a menu where more detailed information can be found, though it is not easy for non-technical users to understand. During the time this study was conducted, the certificate indicator icon distinguished between two security levels: no certificate (the icon reads "Not secure"), and Domain/Organization/Extended Validation (DV/OV/EV; where the icon shows a lock symbol). However, we used an older EV certificate indicator for purposes of a richer security level comparison. This is discussed below in Section 7.1.

Our results suggest that our participants did not primarily base their security-related judgments on the website identity information provided by browsers. Across all tasks in both of our studies, the more visually appealing websites (Hva) were consistently rated as more secure than their less visually appealing (Lva) counterparts. Remarkably, this was true even when the Hva certificate indicator showed "Not secure". The eye-tracking results from Study 1 showed that participants never looked at these indicators as they used the websites, and they also never opened the certificate information menu to see more detailed information. It seems that, in both of our studies, there was a halo effect of visual appeal on security, where participants based their judgments of the website's identity primarily based on its appearance. In other words, we believe we have evidence that, from a user's perspective, *what is beautiful is secure.* In essence, despite any possible education on the meaning of certificates or of their indicators, it appears that users based their judgments of the website's security status on its appearance.

The only situation in which participants appeared to use the certificate indicator to inform their judgments of security was during the choice questions in Study 2. For all choice questions, the top two choices were Hva websites, with

the number one choice being the most secure out of the two HVA options. It appears that participants were using the elimination by aspects heuristic [55], in which one chooses among a number of possibilities by sequentially eliminating some options because they do not possess some attribute of importance. Here, the vast majority of participants seemed to first eliminate two websites based on their visual appeal, and then either (a) choose based on the security level suggested by the certificate indicator, or (b) choose randomly between the two. This pattern of behaviour presents more evidence that our participants primarily make security-relevant decisions online based on the visual appeal of the website. In real usage scenarios, of course, people do not choose among a number of alternative website interfaces, but make decisions based on the single website they are on. Those who appeared to base their decision on the certificate indicators likely did so only because it was the only way of distinguishing between the final two options. We think this finding offers a potential glimmer of hope for website certificate indicators, which we discuss in Section 7.2.

In satisfaction of our secondary research goal, we found that participants in our studies also tended to rate more visually appealing website as more usable, which is consistent with prior work on this topic [53].

7.1 The Trouble with Certificate Indicators in Browser UIs

Some form of the 'legacy' EV indicator used in Study 2 was used by all major web browsers until approximately mid-2019, when it was decided that they would no longer give EV certificates a distinct appearance. In their announcement of this change [6], Chrome developers explain why the decision was made:

> Users do not appear to make secure choices (such as not entering password or credit card information) when the UI is altered or removed, as would be necessary for EV UI to provide meaningful protection. Further, the EV badge takes up valuable screen real estate, can present actively confusing company names in prominent UI, and interferes with Chrome's product direction towards neutral, rather than positive, display for secure connections. (Chromium Security UX Team [6])

In support of the claim of that EV certificates do not positively affect security behaviour, they cite a paper by Google researchers [50]. In one large-scale study, behavioural data was collected from 2% of Chrome users. When the EV indicator was present there was a marginal increase on clicks to the certificate information menu, but no increase in interactions with its functionality. Follow-up Mechanical Turk studies found that the presence of an EV indicator had no effect on participant ratings of how comfortable they would be logging-in to a website. In general, they found that users pay more attention to a website's content than to the browser's UI when making trust decisions. This latter finding is consistent with older studies: for example, users in studies by [13] and [17] use language and spelling to judge if a website is credible. Thompson et al. recommend that browser vendors offer "radical redesigns" of identity indicators in order to improve their effectiveness.

In related recent work, [48] attempted to make certificate information more obvious and accessible by developing a Chrome extension which emphasized basic information about its identity to users. These notifications were straightforward and came from a trustworthy source: the operating system (Windows or macOS). Participants noticed these notifications, but their judgments of website identity, and their responses to post-task interview questions showed that the content of the notifications was ignored. It seemed that users were instead making security-related judgments based on the appearance of the website.

The results of the present study also show users making security-related decisions based website content—in this case it's *appearance*. Browser indicators were seemingly ignored—even the rather explicit "Not secure" signal provided in the absence of a certificate.

In sum, it seems that current certificate indicators (i.e. icons and occasionally text next to the URL) do not provide an effective way for users to distinguish between any of the various certificate levels and the assurances of connection security and identity that they offer. Since users do appear to using website content to inform their security decisions, perhaps one alternative to current web certificate indicators would be to signal certificate information through modification of the website's content instead.

One example of such an approach that is immediately suggested by our results is to make insecure websites look ugly. The results of Study 2 already provide empirical support for this. In the Mismatched condition, the low security websites were shown with high visual appeal, and participants incorrectly rated their security as high. In the Matched condition, the same low security websites were displayed with low visual appeal, and they were correctly rated as low security by participants. In real usage, a browser could additionally offer a message, for example: "This website display is manipulated to emphasize that it is not secure". However, a binary (beautiful/ugly) variable obviously cannot fully convey a website's complex security status, nor can it adequately express most of the identity information supplied by its certificate. The primary purpose of such an approach would be to catch the user's attention, and more information needs to be made available to users (in an understandable format) to help them make informed trust decisions with websites. This uglification technique might help set the stage for the presentation of such information.

7.2 A Glimmer of Hope for Existing Certificate Indicators?

The situation seems bleak for existing browser indicators, but we think one result of our study suggests a possibility for how they might be salvaged.

In the choice questions (see Tables 11 & 12), participants tended to select the highest security option within the high visual appeal conditions as the most secure, the one they would log in to, and so on. Whereas in our other experiments users seemed to ignore the certificate indicators and use the website's appearance to make security-related decisions, in the choice questions they must ultimately choose between two Hva websites that differ only in their certificate indicators. It is unlikely that participants were basing this decision on the full extent of the information that certificate indicators are meant to convey, given that, for example, in the Mismatched condition, they failed to select either of the two Lva websites, which were the most secure. This artificial task forces participants to attend to certificate indicators.

This result suggests that helpful information about website security can be communicated to users if they are required to attend to them. Perhaps the effectiveness of certificate indicators could be improved if browsers made explicit attempts to teach users about their meaning.

Prototypes for educational interventions that temporarily borrow users' attention could be developed and tested to find the best method of portraying website security information. Potential solutions would vary in terms of the amount of information they wish to convey to users, and subsequently the amount of users' time and attention they would take; there is a continuum of target outcomes varying from simple behaviour modification (such as what we observed in Study 2) to improving users' mental models of the concepts relevant to website identity and certificates. In prior work, researchers have developed an interactive tutorial of the latter type as a web app which attempted to teach users about Certificate Authorities, the different types of certificates, and what information they can contain may benefit users by helping them verify the identity of websites [47]. The interactive tutorial was technologically crude, but it was still able to produce significant learning effects. However, the tutorial took several minutes to complete, which would be too demanding in a real-world usage context. Perhaps it is possible that in under one minute enough information could be communicated to users that they are sufficiently aware of (a) the indicators and what they signify, roughly, and (b) references to entities like Certificate Authorities and encryption. The following subsection cites one high-level

approach to designing such interventions. If we made more of an attempt to help users understand the value of website certificate indicators, perhaps their effectiveness could be improved. This possibility could be explored in future work.

7.3 Reducing the Halo Effect of Visual Appeal

There is evidence that greater familiarity with an object helps to reduce halo effects of one of its attributes on another [19, 59]. This implies that the halo effect is a heuristic which is used to aid decision-making in cases where one's knowledge is lacking, and which is relied upon less the more one learns about the object. If true, this would mean that efforts to improve users' mental models of website security as was just discussed in Section 7.2 could not only make certificate indicators more useful, but it could also serve as a robust safeguard against the halo effect of visual appeal. To test this hypothesis, future studies could compare the halo effect of visual appeal with and without instruction about the web certificates and the assurances they offer.

One challenge to this approach is the fact that security is an important, but typically secondary task for users (i.e. people generally do not come to their devices to do security) which puts heavy constraints on how much time and attention we can expect users to invest in learning security concepts. Visual appeal information, on the other hand, is highly available for everyone. Perhaps if security information could be presented to users in a similarly available form they would not rely on visual appeal cues from the website itself in their security decision-making. The difficulty is that security matters are technically complex, and which makes gist explanations non-trivial. Users' understanding of software is largely determined by the system image [32], which normally leaves out security issues. One suggestion for adding security information to the system image without overburdening the user is to leverage visualization techniques (which lend themselves to the simple expression of complex ideas), and abstraction hierarchies such as those used in Ecological Interface Design [57] (making elaborating details available on demand). The all-important particulars of an informative yet economical security system image remain unspecified, and we think represent an opportunity for challenging and interesting research.

8 Conclusions

This research contributes to an improved understanding of perceived security in websites, and of the relationship between visual appeal, security, and usability. From these results, we can conclude that security and usability ratings are strongly influenced by visual appeal. For websites, this means that an easy way to gain a user's trust is to increase the appeal of the website. However, this also means that attackers can take the same approach. In particular, where attackers use visually appealing websites, perhaps cloned from genuine websites, users may be influenced by the visual appeal and disregard any indicators—even ones that say "Not secure". This also calls for the education of users on website security and website certificates. The standardization of the presentation of certificate information along with the differentiation of DV, OV, and now EV certificates is an essential part successful website validation that would benefit any user. Future work could examine the effect of browsers making insecure websites less visually appealing, to see if this would deter users from interacting with them. Furthermore, future work could explore the possibility of applying these results to mobile devices, where the certificate and URL interface are often hidden to save screen space.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grant RGPIN-2015-05629, Postdoctoral Fellowship 516586, Alexander Graham Bell Canada Graduate Scholarship 546531.

References

[1] Greg Aaron. 2021. *APWG Phishing Activity Trends Report.* Anti-Phishing Working Group. Retrieved March 21, 2022 from https://docs.apwg.org/reports/apwg_trends_report_q4_2020.pdf

[2] Simon Bell and Peter Komisarczuk. 2020. An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank. In *Proceedings of the Australasian Computer Science Week Multiconference* (Melbourne, Australia). ACM, New York, NY, USA, 1–11.

[3] Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, and Saranga Komanduri. 2011. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Symposium on Security and Privacy (SP '11)* 9, 2 (3 2011), 18–26. https://doi.org/10.1109/MSP.2010.198

[4] John Brooke. 1996. SUS: A quick and dirty usability scale. In *Usability Evaluation in Industry.* Taylor & Francis, London, UK, 189–194.

[5] Jesse Chandler, Cheskie Rosenzweig, Aaron J Moss, Jonathan Robinson, and Leib Litman. 2019. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior research methods* 51, 5 (2019), 2022–2038.

[6] Chromium Security UX Team. 2019. *EV UI Moving to Page Info.* Google. Retrieved April 10, 2022 from https://chromactivelyium.googlesource.com/chromium/src/+/HEAD/docs/security/ev-to-page-info.md

[7] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). L. Erlbaum Associates, Hillsdale, NJ, USA.

[8] D. Cooper, S. Santesson, S. Farell, S. Boeyen, R. Housley, and W. Polk. 2008. *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile.* RFC 5280. RFC Editor. https://www.rfc-editor.org/info/rfc5280

[9] Dan Craigen, Nadia Diakun-Thibault, and Randy Purse. 2014. Defining Cybersecurity. *Technology Innovation Management Review* 4 (10 2014), 13–21. https://doi.org/10.22215/timreview/835

[10] Karen Dion, Ellen Berscheid, and Elaine Walster. 1972. What is beautiful is good. *Journal of Personality and Social Psychology* 24, 3 (1972), 285.

[11] Susan L. Feagin. 1995. Beauty. In *The Cambridge Dictionary of Philosophy*, Robert Audi (Ed.). Cambridge University Press, Cambridge, UK, 16.

[12] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. 2016. Rethinking Connection Security Indicators. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016).* USENIX Association, Denver, CO, 1–14. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/porter-felt

[13] BJ Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani, et al. 2001. What makes web sites credible? A report on a large quantitative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 61–68.

[14] Karl Grammer, Bernhard Fink, Anders P Møller, and Randy Thornhill. 2003. Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews* 78, 3 (2003), 385–407.

[15] Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human–Computer Interaction* 19, 4 (2004), 319–349.

[16] ISO. 2018. *ISO 9241-11:2018(en): Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.* Standard. International Organization for Standardization, Geneva, Switzerland.

[17] Markus Jakobsson, Alex Tsow, Ankur Shah, Eli Blevis, and Youn-Kyung Lim. 2007. What instills trust? A qualitative study of phishing. In *International Conference on Financial Cryptography and Data Security.* Springer, New York, NY, USA, 356–361.

[18] Peter Kampstra. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of statistical software* 28, 1 (2008), 1–9.

[19] Barbara Black Koltuv. 1962. Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs: General and Applied* 76, 33 (1962), 1.

[20] Masaaki Kurosu and Kaori Kashimura. 1995. Apparent Usability vs. Inherent Usability: Experimental Analysis on the Determinants of the Apparent Usability. In *Conference Companion on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95).* ACM, New York, NY, USA, 292–293. https://doi.org/10.1145/223355.223680

[21] Talia Lavie and Noam Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies* 60, 3 (2004), 269–298.

[22] Sangwon Lee and Richard J Koubek. 2010. Understanding user preferences based on usability and aesthetics before and after actual use. *Interacting with Computers* 22, 6 (2010), 530–543.

[23] Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 1–30.

[24] Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and Judith Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25, 2 (2006), 115–126.

[25] Leib Litman, Jonathan Robinson, and Tzvi Abberbock. 2017. TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods* 49, 2 (2017), 433–442.

[26] Anthony C. Little, Benedict C. Jones, and Lisa M. DeBruine. 2011. Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, 1571 (2011), 1638–1659.

[27] Sascha Mahlke and Gitte Lindgaard. 2007. Emotional Experiences and Quality Perceptions of Interactive Products. In *Human-Computer Interaction. Interaction Design and Usability*, Julie A. Jacko (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 164–173.

[28] Sascha Mahlke and Manfred Thüring. 2007. Studying Antecedents of Emotional Experiences in Interactive Contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07).* ACM, New York, NY, USA, 915–918. https://doi.org/10.1145/1240624.1240762

[29] Morten Moshagen and Meinald Thielsch. 2013. A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology* 32, 12 (2013), 1305–1311.

[30] Net Applications. 2020. *Browser Market Share*. Net Applications. Retrieved July 21, 2020 from https://netmarketshare.com/browser-market-share.aspx

[31] David W. Nordstokke, Bruno D. Zumbo, Sharon L. Cairns, and Donald H. Saklofske. 2011. The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data. *Practical Assessment, Research, and Evaluation* 16, 1 (2011), 5.

[32] Donald A. Norman. 1986. Cognitive Engineering. In *User Centered System Design: New Perspectives on Human-computer Interaction*, Donald A. Norman and Stephen W. Draper (Eds.). CRC Press, Boca Raton, FL, USA, 266–290.

[33] Donald A. Norman. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, New York, NY, USA.

[34] Supavich Fone Pengnate and Rathindra Sarathy. 2017. An Experimental Investigation of the Influence of Website Emotional Design Features on Trust in Unfamiliar Online Vendors. *Computers in Human Behavior* 67 (2017), 49–60.

[35] Nornadiah Mohd Razali and Yap Bee Wah. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics* 2, 1 (2011), 21–33.

[36] Martin Reimann, Judith Zaichkowsky, Carolin Neuhaus, Thomas Bender, and Bernd Weber. 2010. Aesthetic package design: A behavioral, neural, and psychological investigation. *Journal of Consumer Psychology* 20, 4 (2010), 431–441.

[37] Gillian Rhodes. 2006. The evolutionary psychology of facial beauty. *Annual Review of Psychology* 57 (2006), 199–226.

[38] David Robins and Jason Holmes. 2008. Aesthetics and credibility in web site design. *Information Processing & Management* 44, 1 (2008), 386–399.

[39] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. 2007. The Emperor's New Security Indicators. In *2007 IEEE Symposium on Security and Privacy (SP '07)*. IEEE, New York, NY, USA, 51–65.

[40] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.

[41] Shopify Inc. 2016. All Shopify Stores Now Use SSL Encryption Everywhere. https://www.shopify.ca/blog/73511365-all-shopify-stores-now-use-ssl-encryption-everywhere

[42] Pinya Silayoi and Mark Speece. 2004. Packaging and purchase decisions: An exploratory study on the impact of involvement level and time pressure. *British Food Journal* 106, 8 (2004), 607–628.

[43] Devendra Singh and Dorian Singh. 2011. Shape and significance of feminine beauty: An evolutionary perspective. *Sex Roles* 64, 9-10 (2011), 723–731.

[44] Milica Stojmenović. 2016. *Your Reputation Precedes You: The Influence of Expectations on Usability and Visual Appeal in a Web Environment*. Ph. D. Dissertation. Swinburne University of Technology. https://researchbank.swinburne.edu.au/items/e3230536-2ac5-4b92-a11d-8c47b65ab353/1/

[45] Milica Stojmenović and Robert Biddle. 2018. Who Are They? Website Authentication: Certificates and Identity. In *Who Are You?! Adventures in Authentication Workshop (WAY '18)*. 1–5. https://wayworkshop.org/2018/papers/way2018-stojmenovic.pdf

[46] Milica Stojmenovic, John Grundy, Vivienne Farrell, Robert Biddle, and Leonard Hoon. 2016. Does Textual Word-of-Mouth Affect Look and Feel?. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Launceston, Tasmania, Australia) *(OzCHI '16)*. ACM, New York, NY, USA, 257–265. https://doi.org/10.1145/3010915.3010926

[47] Milica Stojmenović, Temitayo Oyelowo, Alisa Tkaczyk, and Robert Biddle. 2018. Building Website Certificate Mental Models. In *Persuasive Technology*, Jaap Ham, Evangelos Karapanos, Plinio P. Morita, and Catherine M. Burns (Eds.). Springer International Publishing, Cham, 242–254. https://doi.org/10.1007/978-3-319-78978-1_20

[48] Milica Stojmenović, Eric Spero, Temitayo Oyelowo, and Robert Biddle. 2019. Website Identity Notification: Testing the Simplest Thing That Could Possibly Work. In *2019 17th International Conference on Privacy, Security and Trust (PST)*. IEEE, New York, NY, USA, 1–7. https://doi.org/10.1109/PST47121.2019.8949048

[49] Omid Taebi, Hamza Aldabbas, and Mary Clarkson. 2013. Users' perception towards usability and aesthetics design of travel websites. In *Proceedings of The International Conference on E-Commerce and Information Technology*, Vol. 117.

[50] Christopher Thompson, Martin Shelton, Emily Stark, Maximilian Walker, Emily Schechter, and Adrienne Porter Felt. 2019. The web's identity crisis: understanding the effectiveness of website identity indicators. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX, Berkeley, CA, USA, 1715–1732.

[51] Manfred Thüring and Sascha Mahlke. 2007. Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology* 42, 4 (2007), 253–264. https://doi.org/10.1080/00207590701396674 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1080/00207590701396674

[52] Noam Tractinsky. 1997. Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 115–122.

[53] Noam Tractinsky, Adi S. Katz, and Dror Ikar. 2000. What is beautiful is usable. *Interacting with Computers* 13, 2 (2000), 127–145. https://doi.org/10.1016/S0953-5438(00)00031-X

[54] Alexandre N. Tuch, Sandra P. Roth, Kasper Hornbæk, Klaus Opwis, and Javier A. Bargas-Avila. 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior* 28, 5 (2012), 1596–1607.

[55] Amos Tversky. 1972. Elimination by aspects: A theory of choice. *Psychological Review* 79, 4 (1972), 281.

[56] Paul Van Schaik and Jonathan Ling. 2009. The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies* 67, 1 (2009), 79–89.

[57] Kim J. Vicente and Jens Rasmussen. 1992. Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics* 22, 4 (7 1992), 589–606. https://doi.org/10.1109/21.156574

[58] WikiMedia Foundation. 2020. *User Agent Breakdowns.* WikiMedia Foundation. Retrieved July 21, 2020 from https://analytics.wikimedia.org/dashboards/browsers/#all-sites-by-browser

[59] Bob T.W. Wu and Susan M. Petroshius. 1987. The halo effect in store image measurement. *Journal of the Academy of Marketing Science* 15, 3 (1987), 44–51.

[60] Min Wu, Robert C. Miller, and Simson L. Garfinkel. 2006. Do security toolbars actually prevent phishing attacks?. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 601–610.

[61] Mel Yamamoto and David R. Lambert. 1994. The impact of product aesthetics on the evaluation of industrial products. *Journal of Product Innovation Management* 11, 4 (1994), 309 – 324. https://doi.org/10.1016/0737-6782(94)90086-8