Int. J. Human–Computer Studies 179 (2023) 103103

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

The effect of prototypicality on webpage aesthetics, usability, and trustworthiness

Aliaksei Miniukovich *, Kathrin Figl

University of Innsbruck, Innrain 52, Innsbruck, 6020, Austria

ARTICLE INFO

Dataset link: https://doi.org/10.7910/DVN/9FKSQI, https://doi.org/10.7910/DVN/XOI0HI, https://doi.org/10.7910/DVN/Z7KLIH

Keywords:
Web design
Prototypicality
First impression
User preferences

ABSTRACT

The user expects webpages of specific categories to have a look-and-feel specific to that category. For example, unlike university homepages, online-shop webpages typically feature relatively little text, a long grid-like structure listing products, and numerous functional elements for product search, filtering, and recommendation. Ensuring that a webpage meets user expectations makes it highly prototypical and improves the user impression of the webpage. Despite the potential impact on users, the concept of webpage prototypicality has not been fully explored or extensively employed in Human-Computer Interaction (HCI). This paper addresses this gap by conducting a user study with 1530 participants to investigate webpage prototypicality. The study revealed a strong correlation between prototypicality and webpage visual aesthetics, perceived pre-use usability, and trustworthiness. Notably, the direct effect of prototypicality on trustworthiness outweighed the indirect effects through aesthetics and usability. Overall, prototypicality, aesthetics, and usability collectively accounted for 29% to 68% of the variance in trustworthiness, depending on a webpage category. These findings underscore the importance of embracing prototypicality within the field of HCI, encouraging its wider adoption.

1. Introduction

Webpage prototypicality is a property of webpage design that describes how closely a webpage resembles the abstract webpage that the user envisions when thinking of a specific category of webpages, such as the webpages of commercial banks or higher education institutions. While webpage prototypicality may appear to be a vague concept, challenging to express through concrete design features (e.g., specific content grid proportions, font sizes, or main and complementary colors) that a prototypical webpage would or would not possess, an examination of highly prototypical webpages often reveals shared visual features among them, resulting in similar appearance. On the other hand, low prototypicality webpages typically exhibit unique differences from both each other and those with high prototypicality. These shared features establish the prototype – the most prototypical webpage – for a particular web category or domain. For example, for the category of higher-education institutions, a prototypical webpage might include a prominent page-wide image at the top, text frequently accompanied by images, a regular content arrangement, a light-colored background, and medium-sized fonts (Fig. 1).

Few HCI studies with webpages have explicitly included prototypicality, e.g., linking it to aesthetics (Tuch et al., 2012a), and it has not been widely adopted in HCI as a useful predictor of user first impression, subsequent preference, and attitude towards websites and the organizations behind them. However, research in cognitive sciences has demonstrated that prototypicality is measurable and influential, impacting the feeling of familiarity and liking for everyday objects (Whitfield, 1983; Monin, 2003), which webpages can be classified as. This paper explores webpage prototypicality and presents a user study that establishes its connection to three key concepts of interest to HCI: webpage trustworthiness, visual aesthetics, and perceived usability.

The study involved 1530 participants and examined over three thousand webpages from three business domains: commercial banking, universities, and ecommerce. The results revealed a strong and linear effect of prototypicality on webpage trustworthiness, which is a novel finding. This effect was both direct and indirect, mediated by visual aesthetics and perceived pre-use usability, though the indirect effect was relatively modest compared to the direct effect. The inclusion of a large sample of participants and stimuli enhances the confidence in the study findings. We propose that webpage prototypicality be utilized in HCI research as a potential predictor of user preferences, and in design practice as a concept for describing and communicating design ideas.

* Corresponding author.
E-mail address: miniukovich@e.atw-lab.com (A. Miniukovich).

https://doi.org/10.1016/j.ijhcs.2023.103103
Received 16 May 2023; Received in revised form 30 June 2023; Accepted 1 July 2023
Available online 10 July 2023

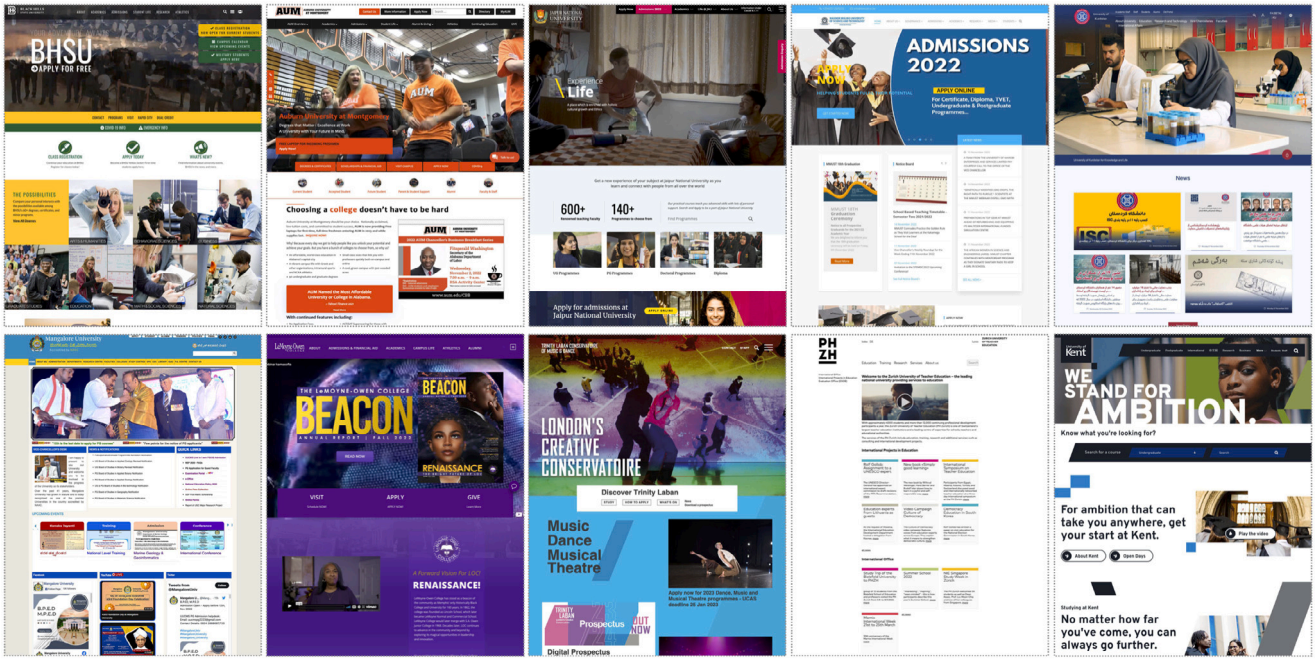

Fig. 1. A selection of homepages of higher education organizations, showcasing examples of high prototypicality (top row) and low prototypicality (bottom row). The selection was based on the user scores of prototypicality, which were adjusted for webpage aesthetics and perceived usability.

2. Theoretical background

When users encounter a website, they often lack specific information about the website or organization behind it to make a decision whether they want to interact with it further or should explore other available alternatives. In such cases, their first impression plays a crucial role in determining whether they choose to stay on the website or leave.

2.1. First impression

First impression refers to the near-instantaneous evaluative judgment of website quality, primarily reflecting users' immediate emotional response upon seeing it. Such an impression forms quickly and remains relatively stable (Lindgaard et al., 2006; Tractinsky et al., 2006), evolving from an immediate-first impression to a deliberate-first impression, and then progressing into the user approach-avoidance tendencies towards the website, with only limited subsequent changes observed (Thielsch et al., 2014; Strebe, 2016). These findings parallel well-established results from evolutionary psychology regarding first impressions of faces (Willis and Todorov, 2006). While the usage after the first impression can modify the perceptions and judgments of website quality (e.g., as when post-use usability influences the judgment of visual aesthetics, known as the pragmatic halo effect, Minge and Thüring (2018)), the original liking-based impression continues to exert influence (Thielsch et al., 2014) and affect users' subconscious choices, even when consciously labeled as irrelevant to the choice-making process (Diefenbach and Hassenzahl, 2011).

First impression (Fig. 2) arises from the perception of visual properties of a webpage, such as visual complexity and aesthetics (Tuch et al., 2012a; Miniukovich and De Angeli, 2014b), likely influenced by the mental effort required to comprehend the webpage (Reber et al., 2004). This initial impression then impacts the judgments of website quality, including expected usability and trustworthiness, which are more cognitively demanding and occur later than the purely liking-based judgments of visual aesthetics (Lindgaard et al., 2011; Skulmowski et al., 2016). However, these quality judgments may not necessarily align with the actual website quality. For example, pre-use perceived usability may not closely correspond to the actual usability (Thielsch et al., 2015), while visual aesthetics – even if it improves users' emotional state and motivation – might not affect the actual user performance on typical tasks (Thielsch et al., 2019). Nevertheless, they both still strongly influence the approach-avoidance behavior towards both the website and the organization associated with it (Ye et al., 2020), potentially driven by a greater inclination to explore websites that make a good impression, particularly in the goal mode (Iten et al., 2018). (In the goal usage mode, users have predefined tasks, while the action mode involves free-form browsing without a specific task.) Behaviors that are impacted include engagement with a website (such as dwell time, fraction of viewed content, and number of visited pages (Strebe, 2016), intention to use a website (Pengnate et al., 2019), intention to revisit or recommend a website to others (Thielsch et al., 2014), preference for a company as a potential employer (Braddy et al., 2008), and willingness to trust and donate to a non-profit organization (Küchler et al., 2020).

2.2. Web design features

Multiple visual features of webpages have been demonstrated to influence first impression, with some being relatively specific and others more abstract. The former include the predominant colors used on the webpage (Reinecke et al., 2013), color saturation (Skulmowski et al., 2016), the number of pictures on the webpage, and the ratios of pictures to text (Douneva et al., 2016), while the latter include visual complexity (Miniukovich and De Angeli, 2014b), design and feature complexity (King et al., 2020), design orderliness (Deng and Poole, 2010; Bakaev and Pogorelova, 2021), website content and information quality (De Angeli et al., 2006), or website informativeness, inspiration, involvement and reciprocity (Kim and Fesenmaier, 2008). However, by far the most studied and cited factor of first impression is the higher-level factor of visual aesthetics (Thielsch et al., 2014; Pengnate et al., 2019; Reinecke et al., 2013; Miniukovich and De Angeli, 2016), with perceived webpage usability and trustworthiness also frequently mentioned (Lindgaard et al., 2011; Thielsch et al., 2014; Skulmowski et al., 2016).

Visual aesthetics is a property of webpage appearance that evokes an immediate visceral pleasant feeling upon perceiving the webpage

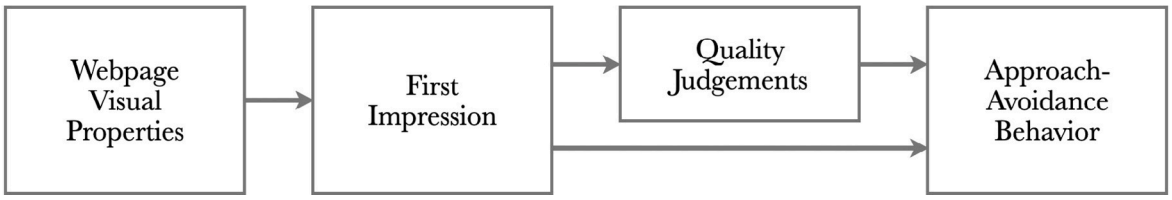

Fig. 2. A schematic representation of first impression, and its antecedents and outcomes, before usage.

(Tractinsky, 2013; Moshagen and Thielsch, 2010; Miniukovich and Marchese, 2020). It dominates the formation of initial impression and user enjoyment of the system (although its effects on user experience diminish over time, Sonderegger et al., 2012; Iten et al., 2018, with actual system usability taking precedence in post-use stages, Saadé and Otrakji, 2007). The appreciation of aesthetics then translates into beliefs and judgments regarding the expected usability and trustworthiness of the webpage (Lindgaard et al., 2011; Skulmowski et al., 2016), potentially due to the halo effect (Hartmann et al., 2008; also see the reverse usability-to-aesthetics halo effect post-use, Minge and Thüring, 2018). Trustworthiness holds particular practical significance for website owners, as they try to establish trust between users and the website. Trustworthiness and trust may encompass several sub-dimensions (Gefen, 2002; Seckler et al., 2015), but they have also often been studied as a unidimensional umbrella judgment (Robins and Holmes, 2008; Lindgaard et al., 2011).

The relationships among aesthetics, usability, and trustworthiness have been well-documented. The perception of aesthetics has been shown to affect expectations of system usability and, in turn, be affected by the actual usability (Tractinsky et al., 2000; Hartmann et al., 2008; Minge and Thüring, 2018). (We use the term "usability" instead of "ease-of-use" because it is much more common, despite its issues (Tractinsky, 2018)). System usability affects system trustworthiness directly (Flavián et al., 2006; Salanitri et al., 2015), and potentially indirectly, through perceived usefulness (Amin et al., 2014), or it specifically affects only the ability sub-dimension of trustworthiness (Casaló et al., 2010). Notably, some studies have proposed a reverse causal link, suggesting that trust influences usability, rather than usability influencing trust (McCloskey, 2006; Hallegatte and Nantel, 2006). Aesthetics has also been found to impact trustworthiness either directly (Robins and Holmes, 2008) or indirectly, through perceived ease of use and usefulness (Li and Yeh, 2010). Several studies have examined all three concepts together (Lindgaard et al., 2011), sometimes implying a temporal transition from aesthetics to usability to trust (Skulmowski et al., 2016), or observing the effect of usability on trust to diminish or be small when aesthetics is accounted for (Oyibo and Vassileva, 2017; Koranteng et al., 2022).

2.3. Prototypicality

Prototypicality is a property that describes how well an object represents its category. For example, English speakers consider a robin to be a more prototypical bird than a penguin, as it better exemplifies the bird category (Hage and Miller, 1976). The concept of prototypicality originates from the prototype theory, which was developed in the cognitive sciences field (Rosch and Mervis, 1975; MacLaury, 1991; Geeraerts, 1989). The theory postulates that the mind relies on categorization to reduce the complexity of the world, and that category membership is graded, not binary (either a member or not), with the object features determining its degree of category belonging. For example, the features "furry," "has paws", and "meows" would be used to define the "cats" category and establish its boundaries. These features also help quantify the extent to which an object belongs to a specific category (Hampton, 1995).

Basic categories (Rosch et al., 1976), such as commercial banking webpages or ecommerce webpages, are likely to exist for webpages and have practical utility. Such categories seem to the user the most natural and convenient for describing differences or similarities between webpages, as superordinate categories might be too generic (e.g., a company homepage), and subordinate categories too narrow to apply in most situations (e.g., commercial banks focusing on wealth management or commercial banks operating only online). While the attempts to develop comprehensive web genre taxonomies have been problematic (Shepherd and Watters, 1998; Santini, 2008; Santini et al., 2011; Crowston et al., 2011) – due to the online genres constantly evolving, overlapping, and being understood differently by different users – they can still be defined situationally for specific purposes. For example, a study on typical webpage element placement (Roth et al., 2010) defined genres such as corporate, social networking, newspaper, ecommerce, search engine, and 'other'.

To be practically useful, such situationally-defined genres should have several properties. They should be large, with many exemplars, so a genre has well-defined boundaries in design space (analyses with few exemplars would only reveal such boundaries with large error margins). They should not be completely dominated by a single exemplar, so the user could judge webpage prototypicality in a study without always directly comparing a webpage against the dominant example, as might be for, e.g., 'google.com' and the search-engine genre. They should be well understood and well known to the average user; otherwise, the user will not be able to judge prototypicality. Finally, they should substantially differ from other genres in their purpose and look-and-feel, to reduce participant confusion about how to classify a particular webpage. The defined web genres would then need to be paired with webpage types (e.g., a homepage, About Us, or Product Listing) for per-webpage prototypicality analyses to be feasible, similar to how the studies on automatic webpage classification would group web genres with webpage types to have relatively clearly defined webpage categories (Kanaris and Stamatatos, 2009).

Once categories are established, prototypicality for the webpages in these categories could be measured using three approaches. First, the older, classical approach (Garrard et al., 2001) would have participants list the features of webpages. The features that are shared by the in-category webpages define that category and could be used to estimate category membership and prototypicality, since the strength of one webpage's category membership is the prototypicality of that webpage. This approach may be impractical for webpages, as it requires many participants who are specifically trained in design to suggest meaningful features. Average users may struggle to identify useful features for webpages, unlike in general knowledge categories such as identifying features like 'ferocious' or 'dangerous' for a lion (Garrard et al., 2001). The second approach relies on the speed of categorization, where higher-prototypicality objects are easier to categorize and categorized faster than lower-prototypicality objects (Mervis and Rosch, 1981). This approach could be used for webpages, but it may be more practical for offline studies, as online participants are often unfocused or lack motivation for reaction-time studies. Finally, the third approach involves participants rating webpages on a prototypicality scale. Some studies used single-item scales (e.g., "this website looks like a typical company website" Tuch et al., 2012a), even though three aspects of prototypicality might require accounting for: exemplar goodness, typicality, and category representativeness (Loken and Ward, 1990).

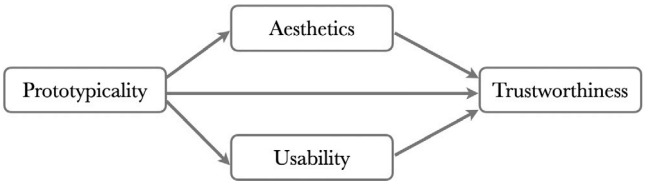

Fig. 3. The expected effects of webpage prototypicality.

Validating the effects of prototypicality on all three aspects of webpage perception, namely aesthetics, usability, and trustworthiness, has only been done for aesthetics (Tuch et al., 2012a). The link between prototypicality and aesthetics is rooted in the human preference for familiar things (Winkielman et al., 2006), at least for the medium levels of familiarity (Carr et al., 2017). This preference has been presented as a theory of aesthetics on its own (Whitfield, 1983), or as a part of the processing fluency theory of aesthetics, where prototypicality increases fluency, which in turn increases aesthetics Reber et al., 2004). The prototypicality-usability link seems likely, as users take longer to find atypically placed webpage objects (Roth et al., 2013; McCarthy et al., 2004). This difficulty in locating objects could lead users to perceive the webpage as having lower usability. The prototypicality-trustworthiness link also seems likely, consistent with the prototypicality-related concept of familiarity affecting risk perception (Song and Schwarz, 2009) and the prototypicality-dependent concept of processing fluency affecting truth/falsehood judgments (Reber and Schwarz, 1999). Fig. 3 summarizes the expected effects of prototypicality.

3. Method

We conducted a study to test if users had a shared idea of a prototypical webpage for three web domains (banking, ecommerce, and higher education) and if prototypicality affected the three major concepts of interest to HCI: webpage visual aesthetics, perceived usability, and trustworthiness.

3.1. Stimuli

We sampled screenshots of homepages from three different web domains: commercial banks, ecommerce shops, and universities. To compile a list of commercial banks, we retrieved URLs from open sources like Wikipedia and the US government financial institution list, in addition to using a search engine. This process generated a pool of 7564 URLs. From this pool, we randomly selected 1500 bank homepages and manually reviewed them, excluding homepages of central banks, wealth management funds, and insurance institutions, focusing only on commercial banking intended for the average person. This process yielded a final sample of 1032 bank homepages, with approximately two-thirds of them being from the US (Miniukovich and Figl, 2023a). A similar process was followed for universities. We initially compiled a list of 6,759 URLs, which was then reduced down to a sample of 1,059 homepages of tertiary education institutions (Miniukovich and Figl, 2023c). Approximately half of the sample consisted of universities from English-speaking countries (the US, UK, Canada, Australia, New Zealand, Ireland, and South Africa). To avoid a potential bias in webpage user scores, overtly religious institutions were excluded from the sample. Finally, due to the limited number of available websites for ecommerce, we reviewed all found websites selling either apparel or homeware (both domains are well-known to the average user compared to more niche shopping domains like fishing) and obtained the homepage screenshots of 550 apparel and 508 homeware ecommerce websites (Miniukovich and Figl, 2023b).

The screenshots were full-page long and mostly 1440 px wide, and were captured using a specifically developed Firefox browser extension. The webpages were sometimes modified prior to screenshotting to preserve their most natural appearance. The modifications included removing full-page pop-ups and cookie notices (GDPR), freezing animations, moving fixed bottom-screen menus to the bottom of the page (otherwise, they would appear in the middle of a full-page screenshot), and fixing in place scroll-dependent background elements (which prevented large empty areas appearing in a full-page screenshot and made the screenshot look more like the original webpage).

3.2. Procedure

After reading the study description and agreeing to the terms and conditions, online participants completed a demographic questionnaire, 83 trials, and a recognition test for quality control (Fig. 4). Each trial involved a light-gray mask being shown for .75-1.25 s, followed by a screenshot that was automatically scrolled down to ensure that participants saw all parts of the webpage, did not dwell on any specific part, and did not skip to the bottom of the webpage using keyboard shortcuts. The top-screen part of the webpage was shown for three seconds, the next part for two seconds, and the remaining parts for one second each. Participants then rated the webpage on one dimension with a seven-point scale. The dimension was the same for a participant across all webpages rated in the session. Prototypicality was measured using three Likert-type items (exemplar goodness: "*This webpage is a representative example of homepages of university/online-shopping/commercial-bank websites,*" family resemblance: "*This webpage has many visual aspects in common with homepages of other university/online-shopping/commercial-bank websites*", and typically: "*This webpage looks like a typical homepage of a university/online-shopping/commercial-bank website*"). The remaining dimensions were measured using a single semantic-differential item each ("*This webpage looks*") with the anchors "*Ugly/Beautiful*" for aesthetics, "*Difficult to use/Easy to use*" for usability, and "*Not trustworthy/Trustworthy*" for trustworthiness. Visual aesthetics and perceived usability are often measured using single-item scales without additional instructions (cf. Sauer and Sonderegger, 2011), and our participants were also not given any additional instructions on how to rate the webpages, apart from the provided items and anchors in our study. In defining and operationalizing visual aesthetics, we followed the strand of HCI research that considers it a design feature that elicits an immediate, cognitively effortless pleasant experience in an average user during and after perceiving the design (Tractinsky, 2013; Moshagen and Thielsch, 2010; Miniukovich and Marchese, 2020).

Participants were permitted to scroll up and down the page freely after the automatic scrolling and could also navigate back to previously-rated screenshots to change their ratings, though very few did (in total, only 9.54% of participants and only for .41% of trials). Each experimental session had 83 trials, including 68 unique screenshots, three training screenshots (the same for all participants), and twelve screenshots shown twice for quality control (selected randomly from the 68 unique ones). The data collection was conducted separately for each web domain, resulting in four datasets (including the apparel and homeware ecommerce websites as separate domains, thus yielding three domains in total but four datasets). The 68 screenshots were randomly selected while trying to ensure that each screenshot received approximately the same number of ratings for each dimension and in total. The ratings provided by non-compliant participants (those not treating the study seriously, see Section 3.3) were continuously removed from per-screenshot rating counters after reviewing data quality for a participant batch, and the ratings were collected again with a new participant batch. The final study step, a recognition test, had participants view the thumbnails (360 by 525 pixels) of ten seen and ten unseen webpages, and indicate the webpages they had seen.

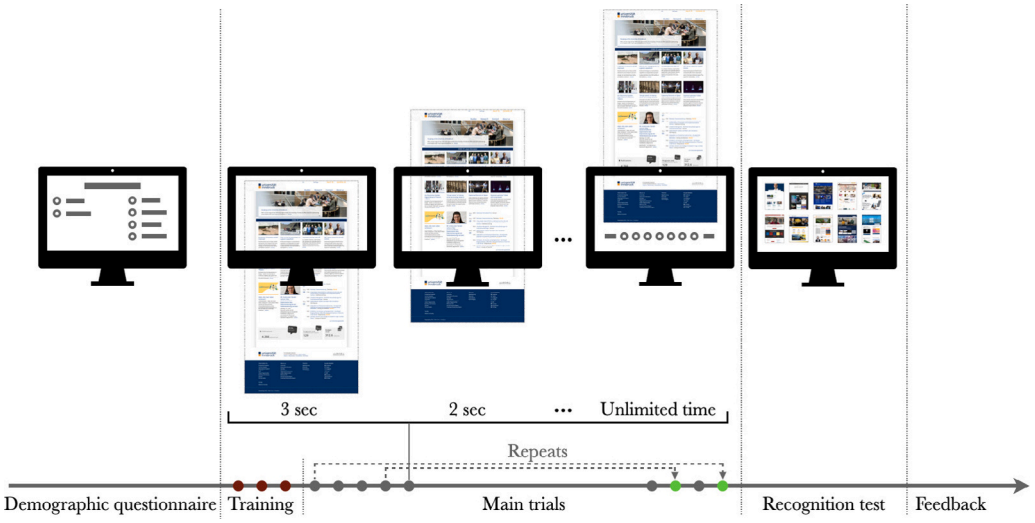

Fig. 4. Study procedure. Each of the 83 trials is represented by a dot on the timeline. The stimuli used for the training trials were consistent across all participants. The twelve stimuli that were shown twice (indicated by green dots) were randomly selected and positioned at least eight trials apart from their corresponding initial-exposure stimuli.

3.3. Data quality assurance

The data of only 1699 out of 3203 participants were initially retained for analyses, confirming a recently observed decline in the quality of crowdsourcing-based data (Chmielewski and Kucker, 2020) due to the emergence of crowdsourcing farms, which are large numbers of organized fake MTurk accounts primarily operating out of South Asia. To distinguish between real and fake participants, we assessed the consistency of the ratings for the twelve twice-shown webpages, the performance of participants on the seen-unseen webpage recognition test, participant demographics (fake accounts often claimed to be 25 years old and from a US state as their origin country, despite individual US states not being in the provided list of countries), usage of VPNs or foreign IP addresses, and optional free-form feedback at the end of the study (as fake accounts tend to write in poor English, frequently use all caps, and provide generic study-unrelated phrases, often repeating the same phrases across multiple accounts). If at least two of these indicators of fake data were unsatisfactory,[1] the participant's data were removed.

The correlations between individual ratings and average ratings were calculated for each of the 1699 participants. A review of the histograms of correlation magnitudes revealed bimodal and negatively skewed normal distributions (Fig. 5), indicating that the data of a number of fake participants still remained in the sample. While there may be inter-individual differences in rating criteria, and not everyone's ratings would strongly correlate with the average (c.f., differences in aesthetics preferences between various demographic groups, Leiva et al., 2022), we expected that the magnitudes of the individual-vs-average score correlations would be normally distributed, especially given the relatively homogeneous participant sample (predominantly USA residents, all fluent English speakers, all active Internet users due to being crowdworkers), and previously demonstrated high inter-rater consistency for related visual features (e.g., webpage visual aesthetics and complexity, Miniukovich and De Angeli, 2014a; and webpage typicality and complexity, Tuch et al., 2012a). To mitigate the impact of unreliable data on the results, we excluded the data of 169 participants from subsequent analyses using the individual-vs-average score correlation coefficients ($r_{with.avg} < .1$) as exclusion criteria. This process resulted in the final sample size of 1530 participants. Fig. 5 depicts the resulting distributions of individual-vs-average correlations.

[1] Except in clear-cut cases, e.g., poor-English feedback replicated across multiple accounts.

3.4. Participants

MTurk crowdworkers were recruited from English-speaking countries (US, UK, AU, CA, NZ, IE). The majority of the 1530 participants (mean age = 40.42 years, SD = 12.53) were from the USA. Among the participants, 857 identified as female, with 8 participants identifying as diverse and 2 participants refusing to answer. 1519 participants reported having normal color vision. Most participants had some higher education (293 were studying, 718 had a BA, 248 had an MSc, and 43 had a Ph.D.), while only 156 participants had not studied after high school (14 had not finished high school), and 72 had technical/vocational training. Participants reported spending an average of 4.68 h per day (SD = 3.13) using the web. They also indicated being relatively familiar with banking (M = 4.06, SD = 1.78, n = 503), apparel ecommerce (M = 5.79, SD = 1.23, n = 290), homeware ecommerce (M = 5.92, SD = 1.14, n = 261), and university websites (M = 3.26, SD = 1.83, n = 476) on a 1–7 scale.

4. Results

We primarily employed structural equation modeling to investigate the direct and indirect effects of prototypicality on trustworthiness, as depicted in Fig. 3, with additional analyses to examine potentially anomalous results, such as an unexpectedly low model fit or seemingly spurious non-linearities in variable relationships.

4.1. Data diagnostics

Some participants used only a sub-scale (e.g., only the ratings 0 to 3), while others used the entire 7-point scale. Due to the limited number of ratings per screenshot (approximately 4.5 data points per screenshot for trustworthiness, usability, and aesthetics, and 7 data points per screenshot for typicality, exemplar goodness, and family resemblance), we could not effectively average out the discrepancies in scale usage by simply calculating the per-screenshot means of non-altered ratings. We instead standardized the ratings within-participant (cf., Tractinsky et al., 2006, see also adjusting scores for response style in Fischer and Milfont, 2010), and then calculated the means.

We then used Mahalanobis distance (alpha = .001) to identify and filter out multivariate outliers (seven for banks, three for universities, seven for apparel, and four for homeware). A review of histograms of per-screenshot means suggested that the data had largely normal distributions for all domains and variables, see Table 1. The three-item measure of prototypicality (family resemblance, typicality, and

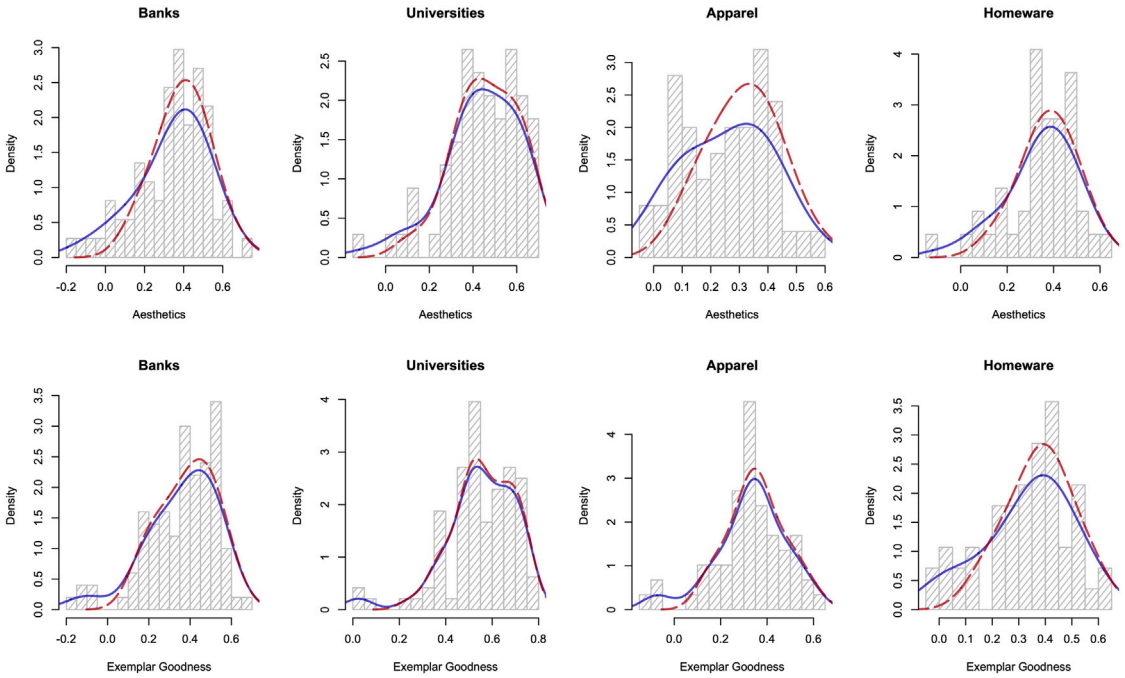

Fig. 5. The distributions of correlation magnitudes between individual and averaged scores for aesthetics (top row) and exemplar goodness (bottom row). The distributions before removing the suspicious no-correlation data are approximated by the blue solid lines, while the distributions after removing the data by the red dashed lines.

Table 1
Descriptive statistics (means, standard deviations, skew, and kurtosis) of the study variables for each of the four datasets.

Variable	Banks			Universities			Apparel			Homeware		
	M(SD)	Skew	Kurt	M(SD)	Skew	Kurt	M(SD)	Skew	Kurt	M(SD)	Skew	Kurt
Aesthetics	−.02(.67)	−.50	.10	−.02(.71)	−.41	−.23	−.01(.61)	−.25	−.02	−.01(.60)	−.50	.06
Usability	.00(.60)	−.79	.48	−.03(.63)	−.56	.14	.03(.61)	−.83	.81	−.03(.68)	−.49	.41
Trustworthiness	−.02(.66)	−.54	−.13	−.03(.71)	−.44	−.32	−.01(.55)	−.33	−.15	−.01(.55)	−.57	.60
Family resemblance	.00(.61)	−.62	.01	−.02(.72)	−.45	−.48	.00(.56)	−.74	.49	−.01(.55)	−.68	.22
Typicality	.01(.57)	−.45	−.21	−.02(.69)	−.53	−.21	.00(.57)	−.57	.46	−.02(.58)	−.51	.05
Exemplar goodness	.00(.58)	−.53	.05	−.03(.72)	−.46	−.30	.00(.54)	−.67	.44	.00(.57)	−.75	.71

Table 2
Cross-correlations among scores for aesthetics, usability (US), trustworthiness (TRU), and prototypicality (PRO) scores for banks (df = 1023) universities (df = 1054), apparel (df = 541), and homeware ecommerce (df = 502). All correlations were significant at $p < .001$. Prototypicality was derived from averaging family resemblance, typicality, and exemplar goodness.

	Banks			Universities			Apparel			Homeware		
	US	TRU	PRO	US	TRU	PRO	US	TRU	PRO	US	TRU	PRO
Aesthetics	.39	.54	.56	.56	.64	.73	.18	.33	.27	.24	.49	.28
Usability	–	.44	.54	–	.49	.61	–	.15	.40	–	.17	.29
Trustworthiness	–	–	.67	–	–	.78	–	–	.45	–	–	.36

exemplar goodness) had a sufficiently high internal consistency for all web domains ($\alpha_{bank} = .84$, $\alpha_{uni} = .92$, $\alpha_{appa} = .82$, $\alpha_{home} = .80$).

4.2. Research model

The means of family resemblance, typicality, and exemplar goodness for each dataset were used as prototypicality, and cross-correlations (Pearson) were calculated for aesthetics, usability, trustworthiness, and prototypicality (Table 2). The correlations were significant for each dataset, as expected, although the magnitude of correlations was larger for commercial banks and universities than for apparel and homeware ecommerce websites.

To explore the direct and indirect effects of prototypicality on trustworthiness, we relied on structural equation modeling, as shown in Fig. 6. The direct effects were large, while the indirect effects via aesthetics were modest, and via usability small and, unexpectedly, negative (Table 3). The residuals of aesthetics and usability correlated, but only for universities and homeware. The model fit was acceptable to high for all web domains, with the exception of homeware. A review of modification indices for homeware revealed that allowing residuals for typicality (not prototypicality, but one of its three items) to correlate with several other indicators could improve model fit. For example, allowing them to correlate with the residuals of family resemblance and exemplar goodness increased the fit to the almost acceptable $RMSEA_{Homeware} = .086$ and Tucker-Lewis Index, $TLI_{Homeware} = .93$. A further review of cross-correlations for homeware (with exemplar goodness, typicality, and family resemblance not averaged in prototypicality, but used as variables on their own) showed typicality to not correlate with aesthetics ($r(502) = .06$, p = n.s.), unlike exemplar goodness ($r(502) = .38$, $p < .001$) and family resemblance ($r(502) = .28$, $p < .001$). Dropping typicality as an indicator from the model for homeware increased model fit ($RMSEA_{Homeware} = .063$; $TLI_{Homeware} = .96$), and this revised model for homeware (with only exemplar goodness and family resemblance) was used for subsequent analyses.

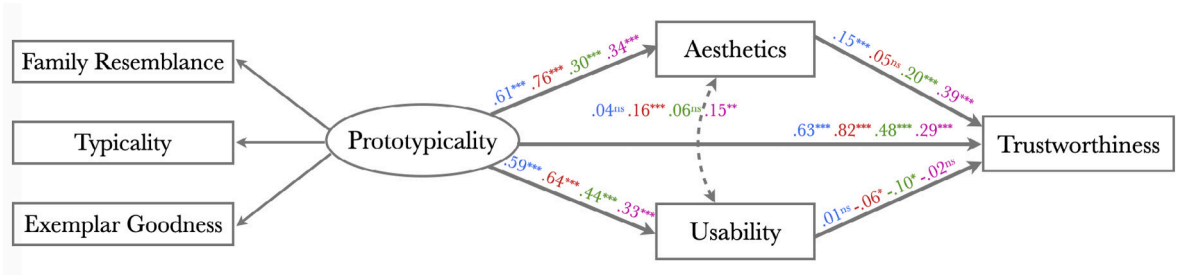

Fig. 6. The tested SEM model. Normalized coefficients are presented from left to right for banks (blue), universities (red), apparel (green), and homeware (purple). Coefficients marked with *** are significant at p < .001, ** at p < .01, and * at p < .05. Non-significant coefficients are marked with ("ns"). The manifest variables of prototypicality were family resemblance, typicality, and exemplar goodness. The residuals of aesthetics and usability were allowed to correlate (shown with a dotted line, with associated coefficients).

Table 3
Results of structural equation modeling for the four datasets (commercial banks, universities, apparel, and homeware ecommerce), with family resemblance, typicality, and exemplar goodness as indicators of the latent factor prototypicality. The residuals of aesthetics and usability were allowed to correlate. [ref] denotes a reference indicator.

		Banks	Universities	Apparel	Homeware
Loadings	Prototypicality → Family resemblance	.81[ref]	.88[ref]	.75[ref]	.78[ref]
	Prototypicality → Typicality	.80***	.87***	.82***	.69***
	Prototypicality → Exemplar goodness	.77***	.90***	.77***	.77***
β scores	Aesthetics ← Prototypicality	.61***	.76***	.30***	.34***
	Usability ← Prototypicality	.59***	.64***	.44***	.33***
	Trustworthiness ← Prototypicality	.63***	.82***	.48***	.29***
	Trustworthiness ← Aesthetics	.15***	.05	.20***	.39***
	Trustworthiness ← Usability	.01	−.06*	−.10*	−.02
	Residuals, Aesthetics ←→ Usability	.04	.16***	.06	.15**
	Indirect effect, Trustworthiness ← Aesthetics ← Prototypicality	.09***	.04	.06***	.13***
	Indirect effect, Trustworthiness ← Usability ← Prototypicality	.01	−.04*	−.04*	−.01
	Total effect, Trustworthiness ← Prototypicality	.73***	.82***	.50***	.42***
R^2	Aesthetics	.37	.58	.09	.12
	Usability	.35	.40	.19	.11
	Trustworthiness	.55	.68	.29	.31
Model Fit	χ^2(df)	7.98(6)	16.93(6)**	24.71(6)***	77.16(6)***
	Tucker-Lewis Index	1.00	.99	.95	.78
	RMSEA	.018	.042	.076	.153

Table 4
Correlations among adjusted usability scores$_{adj}$ (residuals of usability regressed on prototypicality and aesthetics), trustworthiness scores, adjusted trustworthiness scores$_{adj}$ (residuals of trustworthiness regressed on page length), and page length.

Correlations between	Banks	Universities	Apparel	Homeware
Usability$_{adj}$ and Trustw.	−.06*	−.08**	−.13**	−.06
Page length and Usability$_{adj}$	−.28***	−.21***	−.32***	−.32***
Page length and Trustw.	.47***	.55***	.33***	.35***
Usability$_{adj}$ and Trustw.$_{adj}$	.07*	.04	−.02	.05

4.3. Additional analyses

To explore the possible causes of the unexpectedly negative relationship between usability and trustworthiness (Fig. 6), we adjusted usability scores for prototypicality and aesthetics, by regressing them on prototypicality and aesthetics, and taking the resulting linear-model residuals as the adjusted usability scores. We used these adjusted scores to select high- and low-usability$_{adj}$ webpages for visual inspection. The inspection did not reveal any apparent explanation for the negative usability-trustworthiness relationship. However, it did reveal that low-usability webpages were noticeably longer than high-usability pages. Subsequent analyses of page length showed it to correlate negatively with usability and positively with trustworthiness. When the effect of page length was accounted for, the usability-trustworthiness relationship became non-significant (Table 4).

To investigate the possible causes of the absence of a correlation between typicality and aesthetics for homeware, we visually examined homeware webpages with high typicality and low aesthetics, and with low typicality and high aesthetics, as these cases may have interfered with the expected positive relationship between typicality and aesthetics. The examination (Fig. 7) revealed that webpages with low typicality and high aesthetics were selling luxury-related products, while those with high typicality and low aesthetics were selling products aimed at the average earner. The typicality-aesthetics relationship might have been significant if either the aesthetics scores were adjusted for the target customer's wealth (e.g., measured using a semantic differential item with anchors "average earner" and "wealthy" in a future study), or if the definition of the homeware ecommerce category accounted for it (using a narrower category).

Data plots revealed the effects of prototypicality on trustworthiness to be linear, and on aesthetics and usability mostly linear, with slight tendencies towards inverted U-shapes (Fig. 8, Table 5). Similarly, the effects of aesthetics and usability on trustworthiness were mostly linear with minor U-shape tendencies, which mostly disappeared when accounting for prototypicality (Table 5, bottom three groups of models). The aesthetics-usability relationship may have been non-linear, featuring an inverted U-shape component for all datasets, except apparel ecommerce, but the contribution of the quadratic component was still modest (Table 5).

5. Interpretation and discussion

Prototypicality strongly affected webpage trustworthiness (Table 3, the total effect row), and aesthetics and perceived usability (Fig. 6). While its effects on webpage aesthetics have already been observed (Tuch et al., 2012a), the effects on perceived webpage usability and trustworthiness are novel findings. These results underscore the importance of webpage prototypicality as a valuable concept for both HCI research

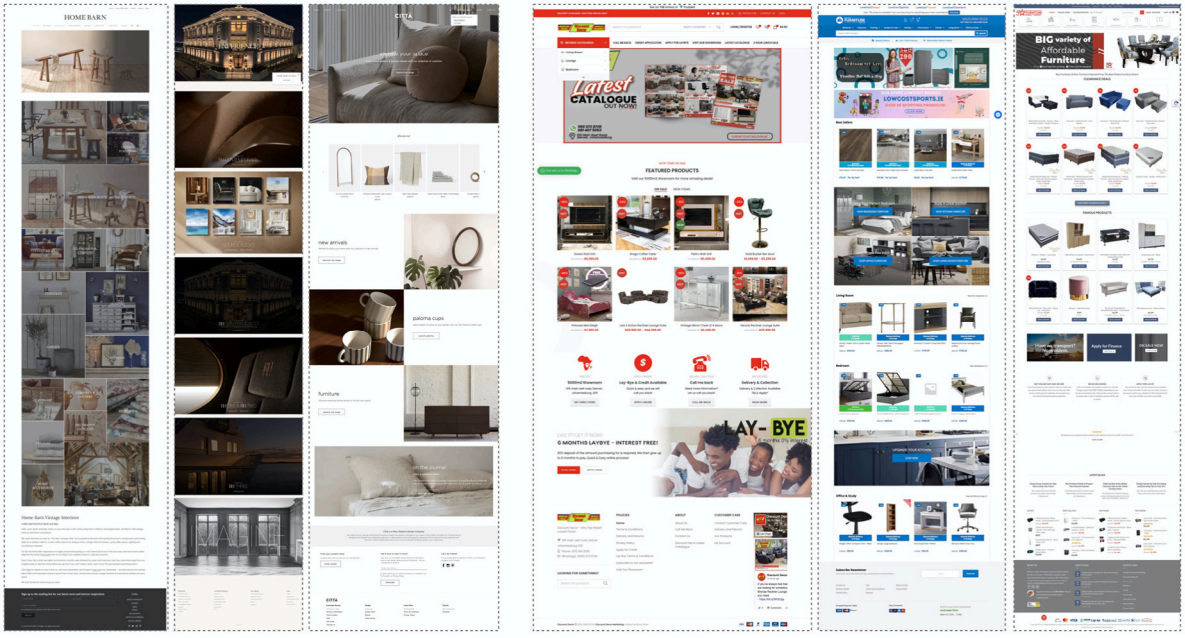

Fig. 7. A selection of homeware webpages showcasing low-typicality-high-aesthetics (left three) and high-typicality-low-aesthetics (right three). These examples illustrate the emphasis of the former on luxury products targeting wealthier customers, while the latter focuses on mass-market products catering to average earners.

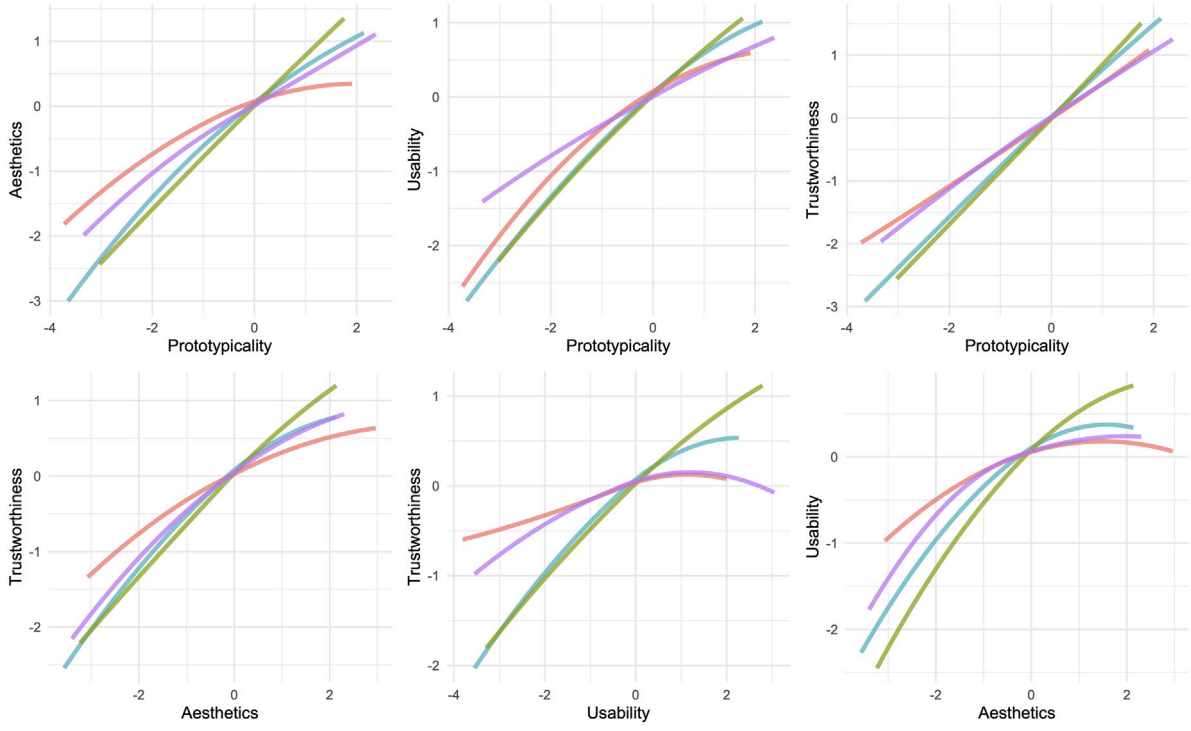

Fig. 8. Smoothed curve plots (locally weighted scatterplot smoothing, LOESS, $\alpha = 1.5$), illustrating potential non-linear relationships among variables (banks - red, universities - purple, apparel - green, homeware - turquoise). Prototypicality scores were derived from the structural-equation models.

and design practice. Future research could use it as a likely predictor of first impressions and related outcomes, such as purchase intention or likelihood of choosing a website from among equivalent alternatives for leisurely browsing or to complete a task. Design practice may want to ensure that their designs do not deviate too much from what the user views as a prototypical webpage for a given business domain, lest the resulting webpages look untrustworthy.

The effects of aesthetics and perceived usability on trustworthiness (Table 2) weakened or disappeared when prototypicality was accounted for (Fig. 6). This finding adds to the existing research on the aesthetics and usability effects on trustworthiness (Lindgaard et al., 2011), suggesting that aesthetics and usability may not be the primary determinants of trustworthiness, but rather partial proxies for prototypicality and its effects on trustworthiness (see also Table 3, the

Table 5
A review of the between-variable relationships for non-linearity was conducted using linear regressions. R^2_{total} shows the total explained variance for the models with both linear and quadratic terms; R^2_{extra} indicates the increase in the explained variance resulting from the addition of a quadratic term. Significance levels are denoted by * p < .05 , * p < .01 and *** p < .001. The bottom three groups of models (with "+ Prototypicality") include prototypicality as an additional linear predictor without displaying its beta coefficients.

Outcome ← Predictor	Domain	Linear β	Quadratic β	R^2_{total}	R^2_{extra}
Aesthetics ← Prototypicality	Banks	.65***	−.07**	.43	.005
	Universities	.79***	−.01	.62	.000
	Apparel	.33***	−.08*	.11	.007
	Homeware	.50***	−.05	.25	.002
Usability ← Prototypicality	Banks	.63***	−.08**	.40	.006
	Universities	.66***	−.04	.43	.002
	Apparel	.48***	−.11**	.24	.012
	Homeware	.38***	−.03	.15	.001
Trustworthiness ← Prototypicality	Banks	.78***	−.01	.61	.000
	Universities	.85***	.00	.72	.000
	Apparel	.55***	.01	.30	.000
	Homeware	.56***	−.01	.31	.000
Trustworthiness ← Aesthetics	Banks	.54***	−.10***	.30	.010
	Universities	.64***	−.03	.42	.001
	Apparel	.33***	−.06	.11	.004
	Homeware	.49***	−.08*	.24	.006
Trustworthiness ← Usability	Banks	.44***	−.09**	.20	.008
	Universities	.49***	−.04	.24	.001
	Apparel	.15***	−.04	.02	.002
	Homeware	.17***	−.07	.03	.005
Usability ← Aesthetics	Banks	.39***	−.13***	.17	.017
	Universities	.56***	−.10***	.33	.011
	Apparel	.18***	−.07	.04	.005
	Homeware	.24***	−.12**	.07	.014
Trustworthiness ← Aesthetics + Prototypicality	Banks	.06*	.00	.61	.000
	Universities	−.07*	.03	.72	.001
	Apparel	.17***	.01	.32	.000
	Homeware	.28***	−.03	.37	.001
Trustworthiness ← Usability + Prototypicality	Banks	−.08**	−.02	.61	.000
	Universities	−.11***	.00	.73	.000
	Apparel	−.14***	.05	.32	.002
	Homeware	−.05	−.03	.31	.001
Usability ← Aesthetics + Prototypicality	Banks	−.04	−.05*	.40	.002
	Universities	.13***	−.06**	.44	.004
	Apparel	.03	.00	.23	.000
	Homeware	.08	−.08	.16	.006

indirect effect rows). They explained only a modest amount of variance in trustworthiness that was not already explained by prototypicality.

The widely discussed relationship between aesthetics and usability (Tractinsky et al., 2000), at least for webpages before use (cf., Minge and Thüring, 2018), appeared to weaken when prototypicality was accounted for (Table 3, the Aesthetics ←→ Usability row). This finding suggests that prototypicality might be a source of the variance shared by both aesthetics and usability, partially explaining the aesthetics-usability relationship. This finding contributes to the discussion of the interplay between aesthetics and usability (Hassenzahl, 2004; Tuch et al., 2012b; Minge and Thüring, 2018).

The strong prototypicality-aesthetics relationship (Table 3, the Aesthetics ← Prototypicality row) supports previous research on webpages (Tuch et al., 2012a) and is consistent with the theories of aesthetics that incorporate prototypicality, such as the preference-for-prototypes theory (Whitfield and Slatter, 1979; Whitfield, 1983) or processing fluency theory (Reber et al., 2004). However, the strong relationship between prototypicality and aesthetics only translated into a modest *indirect* effect of prototypicality on trustworthiness (via aesthetics, Table 3, the indirect effect row for aesthetics). While significant (similar to the past research on package design with purchase intent as an outcome instead of trustworthiness, Celhay and Trinquecoste, 2015), it was substantially weaker than the *direct* effect. The strong *direct* prototypicality-trustworthiness effect suggests that first impression – and follow-up outcomes, such as website- and company-related preferences and judgments – may not only depend on aesthetics but also on other visual webpage features. This finding adds to the discussion of the primacy of aesthetics in forming the first impression and subsequent preferences (Thielsch et al., 2014; Lee and Koubek, 2012).

When adjusted for prototypicality and aesthetics, perceived usability unexpectedly tended to negatively affect trustworthiness (Table 3, the Trustworthiness ← Usability row; and Table 4, the $Usability_{adj}$ and Trustw. row). We could not identify a satisfactory explanation for the negative effect (such as, an explicit bias in our webpage sample, or a perception- and design-related correlate that reversed the relationship from positive to negative), but we did observe low-usability (usability adjusted for aesthetics and prototypicality) webpages to be considerably longer than high-usability webpages. Webpage length correlated negatively with perceived usability and positively with trustworthiness – longer webpages were seen as difficult to use, but trustworthy – and when trustworthiness was adjusted for page length, usability no longer tended to correlate negatively with it (Table 4, last row). Future research should determine whether page length itself causes cross-directional effects on usability and trust, or whether it simply manifests another, design-relevant confounding variable that we failed to recognize.

The three-item metric of webpage prototypicality, comprising of family resemblance, typicality, and exemplar goodness, seemed to be a satisfactory three-item metric of webpage prototypicality, with relatively high internal consistency (Cronbach's α from .80 to .92). Participants also did not report any substantial difficulties in using the three items in their post-study feedback. Their only notable concern was the limited variance in aesthetics observed in the ecommerce samples, as they perceived the majority of ecommerce webpages to be visually

appealing. However, the SEM for homeware ecommerce webpages did have a relatively low fit, which one of the indicators – typicality – appeared to cause. Visually inspecting the homeware webpage sample (Fig. 7) suggested that the homeware ecommerce category might have needed to be split into two sub-categories based on the user income: one for luxury shoppers and one for average earners.

The relationship between trustworthiness and prototypicality was linear, while the aesthetics-prototypicality and usability-prototypicality relationships were largely linear, with a small inverted U-shape component (at least for banking and apparel ecommerce, as shown in Fig. 8, and Table 5). This suggests that increasing prototypicality beyond a certain point did not add to webpage aesthetics and perceived usability. We could speculate that webpages with very high prototypicality might look slightly clichéd to the user, lowering aesthetics, and outdated, lowering usability.

The findings across the four webpage categories (banking, universities, apparel, and homeware) did not contradict each other. The only notable difference between the categories was the strength of between-variable relationships being generally weaker for ecommerce (both apparel and homeware) than for banking and universities (e.g., see Table 2 left half versus right half). The difference might be explained by the relative homogeneity of the ecommerce sample—most ecommerce pages seemed very well designed, with teams of professionals seemingly updating them frequently to reflect latest trends. Future research may need to avoid studying novel phenomena using samples of actual, real-world ecommerce webpages that have not been manipulated for experimental purposes.

6. Methodological implications

Ensuring sufficient quality of our data was crucial for confidence in study findings, but this was challenging and laborious due to the numerous "farmer" MTurk accounts (purchased or stolen MTurk accounts of mainly US residents, "farming" research surveys for income) submitting nonsensical responses (e.g., the same rating for all webpages, or random ratings). Future research could adopt our approach to handling fake data, which involved using several non-trivial indicators at the same time to detect and filter out nonsensical responses. The indicators included a recognition test, rate-rerate consistency test, free-form feedback analysis, demographic information review, and several other indicators.

Even despite using these more sophisticated indicators than the often-used trivial attention checks (such as a multiple-choice question explicitly asking a participant to choose an option, e.g., option "c," Barends and De Vries, 2019), some fake accounts still went undetected, and their data had to be filtered out during the data analysis step (see Fig. 5). MTurk's native worker-quality metrics, such as the number of completed tasks and acceptance rate, could not help, as fake accounts often had thousands of completed tasks with > 99% acceptance rate, possibly due to successfully "completing" self-created tasks, which can be created very cheaply. Relying only on the detection of VPN use (a common technique, Kennedy et al., 2020) would not help either, as many real workers used VPNs, while many fake workers had IPs that were not detected as coming from VPNs.

To solve the issue of fake workers, we suggest assembling and maintaining lists of "bad" MTurk accounts, previously detected as farming accounts, and "good" accounts, which have completed surveys and passed all quality checks.[2] One approach could involve admitting all accounts to studies except those in the "bad" list, while the other could involve admitting only the accounts from the "good" list. Such lists already exist (Hauser et al., 2022), but they are part of commercial solutions and may need to be re-implemented as free research infrastructure[3] to reduce the strain on public funds supporting basic research.

7. Limitations

The limitations of this study include having a viewing-only procedure, limited number of design-related constructs, significant amount of fraudulent participant data, relatively homogeneous participant sample, and only desktop webpages as stimuli. Future research should have an experimental procedure with a scenario and actual website usage instead of just screenshot viewing, as this may lead to partially different results and conclusions. For example, the effect of usability on trustworthiness may become stronger after actual usage. A corresponding study with live websites would have realistic tasks for higher ecological validity, and separation of tasks in goal- and action-oriented as a between-subjects factor, to test for the effects of usage mode (cf., Iten et al., 2018). The stimuli would be relatively few websites – impractical to have many since real usage takes much longer than just viewing – pre-selected in high- and low-prototypicality groups (a within-subjects factor). The pre-selection could involve first rating the prototypicality of a larger group of websites in a separate pre-study (no-usage procedure to allow for rapid larger-volume evaluation) and then sampling the high- and low-prototypicality groups from the extremes of the resulting prototypicality rating distribution. Due to the longer usage procedure, each participant would rate stimuli on all design dimensions, not just one.

Future studies could also include additional design-related constructs, such as webpage perceived usefulness or visual complexity, to further clarify how the perception of webpage trustworthiness is formed and what other factors might affect it. The studies might further benefit from recruiting participants who are motivated not solely by monetary incentives and provide thoughtful responses. This would allow for a reliable estimation of inter-rated consistency for the measured constructs rather than, as in our study, relying on consistency to identify fraudulent participation.

Future research could investigate whether the results generalize to participants who do not speak English, and whether any demographic variables moderate the relationships shown in Fig. 6. We would not expect the moderation to be large, as demographics more likely affect what is viewed as prototypical (the emblematic, typical, average look for a webpage category) rather than the effects of prototypicality itself on other variables. Moreover, what is viewed as prototypical might only modestly depend on demographics, e.g., while gender might affect preference for specific design aspects (Tuch et al., 2010; Lin and Hsieh, 2016), websites are very rarely customized for a gender, and everybody sees the same designs and learns to see the same designs as prototypical. Age could affect prototypicality, as older users likely encountered older, currently-outdated designs, and such encounters would accumulate into a more old-fashioned view of what is prototypical. However, we would expect such an older-design bias to be small, due to the fresh encounters with current design relatively quickly overriding older encounters. Finally, country of origin and cultural background might affect prototypicality more substantially than either gender or age. While average-user websites (e.g., not those dedicated to niche local fashion or arts) seemingly rely more and more on the global design ideas and trends, local brands may still want to visually stand out and connect with the local user by relying on a local design tradition and local user expectations.

The study only included desktop websites, and future studies may need to include mobile websites and apps. We would, however, expect that our conclusions generalize across device types and screen sizes.

[2] Our lists with several thousand worker IDs can be found here: https://github.com/aliko-str/mturk.worker.lists

[3] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/research-infrastructures_en

While mobile websites do differ from desktop websites in their look-and-feel, we would still expect them to have a prototype - a typical, emblematic look for a web category and specific to mobile websites. Having a prototype allows for the phenomenon of prototypicality (the distance from an exemplar to the prototype), and the effects of prototypicality on first impression, described in this paper. We might expect these effects to be smaller than for desktop websites due to a more limited design space of mobile devices, e.g., because of a smaller screen or several very widely adopted mobile-design ideas, such as 'hamburger' menus. These mobile-specific characteristics could limit the potential visual differences among mobile websites, thus limiting the variance in design-related variables, which would lower the chances for covariance among them and weaken potential effects, including the effects of prototypicality.

8. Conclusion

This paper introduced the concept of webpage prototypicality and explored it in a user study. The study found that prototypicality strongly affected the user perception of webpage visual aesthetics, usability, and trustworthiness. The effects on trustworthiness were direct and indirect, partially mediated by aesthetics and usability. Aesthetics and usability themselves also affected trustworthiness, but their unique, prototypicality-unrelated effect was relatively small. These results were replicated across four large webpage datasets from three web domains (commercial banking, universities, and apparel and homeware ecommerce), increasing confidence in their validity. Future research should include prototypicality in studies and theoretical models of user preference and decision-making, as it appears to be a measurable and influential construct.

CRediT authorship contribution statement

Aliaksei Miniukovich: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. **Kathrin Figl:** Conceptualization, Methodology, Writing – original draft, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Aliaksei Miniukovich reports financial support was provided by Austrian Science Fund.

Data availability

Data available at:
https://doi.org/10.7910/DVN/9FKSQI,
https://doi.org/10.7910/DVN/XOI0HI and
https://doi.org/10.7910/DVN/Z7KLIH.

Acknowledgments

This research was funded in whole, or in part, by the Austrian Science Fund, Austria (FWF), project grant M-2827.

References

Amin, M., Rezaei, S., Abolghasemi, M., 2014. User satisfaction with mobile websites: the impact of perceived usefulness (PU), perceived ease of use (PEOU) and trust. Nankai Bus. Rev. Int. 5 (3), 258–274. http://dx.doi.org/10.1108/NBRI-01-2014-0005.

Bakaev, M.A., Pogorelova, A.O., 2021. Profiling of website visitors based on dimensions of user experience. In: XV International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering. APEIE, IEEE, pp. 433–438.

Barends, A.J., De Vries, R.E., 2019. Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. Pers. Ind. Differ. 143, 84–89.

Braddy, P.W., Meade, A.W., Kroustalis, C.M., 2008. Online recruiting: The effects of organizational familiarity, website usability, and website attractiveness on viewers' impressions of organizations. Comput. Hum. Behav. 24 (6), 2992–3001.

Carr, E.W., Huber, D.E., Pecher, D., Zeelenberg, R., Halberstadt, J., Winkielman, P., 2017. The ugliness-in-averageness effect: Tempering the warm glow of familiarity. J. Personal. Soc. Psychol. 112 (6), 787–812.

Casaló, L.V., Flavián, C., Guinalíu, M., 2010. Generating trust and satisfaction in e-services: the impact of usability on consumer behavior. J. Rel. Mark. 9 (4), 247–263.

Celhay, F., Trinquecoste, J.F., 2015. Package graphic design: Investigating the variables that moderate consumer response to atypical designs. J. Prod. Innov. Manage. 32 (6), 1014–1032.

Chmielewski, M., Kucker, S.C., 2020. An MTurk crisis? Shifts in data quality and the impact on study results. Soc. Psychol. Pers. Sci. 11 (4), 464–473.

Crowston, K., Kwaśnik, B., Rubleske, J., 2011. Problems in the use-centered development of a taxonomy of web genres. In: Mehler, A., Sharoff, S., Santini, M. (Eds.), Genres on the Web: Computational Models and Empirical Studies. Springer Netherlands, Dordrecht, pp. 69–84. http://dx.doi.org/10.1007/978-90-481-9178-9_4.

De Angeli, A., Sutcliffe, A., Hartmann, J., 2006. Interaction, usability and aesthetics: What influences users' preferences? In: Proceedings of the 6th Conference on Designing Interactive Systems. DIS '06, Association for Computing Machinery, New York, NY, USA, pp. 271–280. http://dx.doi.org/10.1145/1142405.1142446.

Deng, L., Poole, M.S., 2010. Affect in web interfaces: A study of the impacts of web page visual complexity and order. MIS Q. 34 (4), 711–730. http://dx.doi.org/10.2307/25750702.

Diefenbach, S., Hassenzahl, M., 2011. The dilemma of the hedonic - Appreciated, but hard to justify. Interact. Comput. 23 (5), 461–472.

Douneva, M., Jaron, R., Thielsch, M.T., 2016. Effects of different website designs on first impressions, aesthetic judgements and memory performance after short presentation. Interact. Comput. 28 (4), 552–567.

Fischer, R., Milfont, T.L., 2010. Standardization in psychological research. Int. J. Psychol. Res. 3 (1), 88–96.

Flavián, C., Guinalíu, M., Gurrea, R., 2006. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Inform. Manag. 43 (1), 1–14.

Garrard, P., Lambon Ralph, M.A., Hodges, J.R., Patterson, K., 2001. Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. Cogn. Neuropsychol. 18 (2), 125–174.

Geeraerts, D., 1989. Introduction: Prospects and problems of prototype theory. Linguistics 27 (4), 587–612.

Gefen, D., 2002. Reflections on the dimensions of trust and trustworthiness among online consumers. In: ACM SIGMIS Database: The DATABASE for Advances in Information Systems. Vol. 33. No. 3. ACM, New York, NY, USA, pp. 38–53.

Hage, P., Miller, W.R., 1976. 'Eagle'='bird': a note on the structure and evolution of shoshoni ethnoornithological nomenclature 1. Am. Ethnol. 3 (3), 481–488.

Hallegatte, D., Nantel, J., 2006. The intertwined effect of perceived usefulness, perceived ease of use and trust in a website on the intention to return. E-Bus. Rev. 6, 1–5.

Hampton, J.A., 1995. Testing the prototype theory of concepts. J. Memory Lang. 34 (5), 686–708.

Hartmann, J., Sutcliffe, A., Angeli, A.D., 2008. Towards a theory of user judgment of aesthetics and user interface quality. ACM Trans. Comput.-Hum. Interact. 15 (4), 1–30.

Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. Hum.–Comput. Interact. 19 (4), 319–349.

Hauser, D.J., Moss, A.J., Rosenzweig, C., Jaffe, S.N., Robinson, J., Litman, L., 2022. Evaluating CloudResearch's approved group as a solution for problematic data quality on MTurk. Behav. Res. Methods http://dx.doi.org/10.3758/s13428-022-01999-x.

Iten, G.H., Troendle, A., Opwis, K., 2018. Aesthetics in context—the role of aesthetics and usage mode for a website's success. Interact. Comput. 30 (2), 133–149.

Kanaris, I., Stamatatos, E., 2009. Learning to recognize webpage genres. Inf. Process. Manage. 45 (5), 499–512.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P.D., Jewell, R., Winter, N.J., 2020. The shape of and solutions to the MTurk quality crisis. Political Sci. Res. Methods 8 (4), 614–629.

Kim, H., Fesenmaier, D.R., 2008. Persuasive design of destination web sites: An analysis of first impression. J. Travel Res. 47 (1), 3–13.

King, A.J., Lazard, A.J., White, S.R., 2020. The influence of visual complexity on initial user impressions: Testing the persuasive model of web design. Behav. Inform. Technol. 39 (5), 497–510.

Koranteng, F.N., Ham, J., Wiafe, I., Matzat, U., 2022. The role of usability, aesthetics, usefulness and primary task support in predicting the perceived credibility of academic social networking sites. Behav. Inform. Technol. 41 (16), 3617–3632.

Küchler, L., Hertel, G., Thielsch, M.T., 2020. Are you willing to donate? relationship between perceived website design, trust and donation decisions online. In: Proceedings of Mensch Und Computer. pp. 223–227.

Lee, S., Koubek, R.J., 2012. Users' perceptions of usability and aesthetics as criteria of pre-and post-use preferences. Eur. J. Ind. Eng. 6 (1), 87–117.

Leiva, L.A., Shiripour, M., Oulasvirta, A., 2022. Modeling how different user groups perceive webpage aesthetics. Universal Access in the Information Society, http://dx.doi.org/10.1007/s10209-022-00910-x.

Li, Y.-M., Yeh, Y.-S., 2010. Increasing trust in mobile commerce through design aesthetics. Comput. Hum. Behav. 26 (4), 673–684.

Lin, C.J., Hsieh, T.-L., 2016. Exploring the design criteria of website interfaces for gender. Int. J. Ind. Ergon. 53, 306–311. http://dx.doi.org/10.1016/j.ergon.2016.02.002.

Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P., 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Trans. Comput.-Hum. Interact. 18 (1), 1–30.

Lindgaard, G., Fernandes, G., Dudek, C., Brown, J., 2006. Attention web designers: You have 50 milliseconds to make a good first impression!. Behav. Inform. Technol. 25 (2), 115–126.

Loken, B., Ward, J., 1990. Alternative approaches to understanding the determinants of typicality. J. Consum. Res. 17 (2), 111–126.

MacLaury, R.E., 1991. Prototypes revisited. Ann. Rev. Anthropol. 20 (1), 55–74.

McCarthy, J.D., Sasse, M.A., Riegelsberger, J., 2004. Could i have the menu please? an eye tracking study of design conventions. In: People and Computers XVII–Designing for Society: Proceedings of HCI 2003. Springer, pp. 401–414.

McCloskey, D.W., 2006. The importance of ease of use, usefulness, and trust to online consumers: An examination of the technology acceptance model with older customers. J. Organ. End User Comput. (JOEUC) 18 (3), 47–65.

Mervis, C.B., Rosch, E., 1981. Categorization of natural objects. Ann. Rev. Psychol. 32 (1), 89–115.

Minge, M., Thüring, M., 2018. Hedonic and pragmatic halo effects at early stages of User Experience. Int. J. Hum.-Comput. Stud. 109, 13–25. http://dx.doi.org/10.1016/j.ijhcs.2017.07.007.

Miniukovich, A., De Angeli, A., 2014a. Quantification of interface visual complexity. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. pp. 153–160.

Miniukovich, A., De Angeli, A., 2014b. Visual impressions of mobile app interfaces. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. pp. 31–40.

Miniukovich, A., De Angeli, A., 2016. Webpage aesthetics: One size doesn't fit all. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. pp. 1–10.

Miniukovich, A., Figl, K., 2023a. Web Design Prototypicality: Commercial Banks. Harvard Dataverse, http://dx.doi.org/10.7910/DVN/Z7KLIH.

Miniukovich, A., Figl, K., 2023b. Web Design Prototypicality: eCommerce. Harvard Dataverse, http://dx.doi.org/10.7910/DVN/9FKSQI.

Miniukovich, A., Figl, K., 2023c. Web Design Prototypicality: Universities. Harvard Dataverse, http://dx.doi.org/10.7910/DVN/XOI0HI.

Miniukovich, A., Marchese, M., 2020. Relationship between visual complexity and aesthetics of webpages. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, New York, NY, USA, pp. 1–13. http://dx.doi.org/10.1145/3313831.3376602.

Monin, B., 2003. The warm glow heuristic: when liking leads to familiarity. J. Personal. Soc. Psychol. 85 (6), 1035.

Moshagen, M., Thielsch, M.T., 2010. Facets of visual aesthetics. Int. J. Hum.-Comput. Stud. 68 (10), 689–709.

Oyibo, K., Vassileva, J., 2017. The interplay of aesthetics, usability and credibility in mobile website design and the effect of gender. J. Interact. Syst. 8 (2).

Pengnate, S., Sarathy, R., Lee, J., 2019. The engagement of website initial aesthetic impressions: an experimental investigation. Int. J. Hum.–Comput. Interact. 35 (16), 1517–1531.

Reber, R., Schwarz, N., 1999. Effects of perceptual fluency on judgments of truth. Conscious. Cogn. 8 (3), 338–342.

Reber, R., Schwarz, N., Winkielman, P., 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? Pers. Soc. Psychol. Rev. 8 (4), 364–382.

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., Gajos, K.Z., 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '13, Association for Computing Machinery, New York, NY, USA, pp. 2049–2058. http://dx.doi.org/10.1145/2470654.2481281.

Robins, D., Holmes, J., 2008. Aesthetics and credibility in web site design. Inf. Process. Manage. 44 (1), 386–399.

Rosch, E., Mervis, C.B., 1975. Family resemblances: Studies in the internal structure of categories. Cogn. Psychol. 7 (4), 573–605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. Cogn. Psychol. 8 (3), 382–439.

Roth, S.P., Schmutz, P., Pauwels, S.L., Bargas-Avila, J.A., Opwis, K., 2010. Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? Interact. Comput. 22 (2), 140–152.

Roth, S.P., Tuch, A.N., Mekler, E.D., Bargas-Avila, J.A., Opwis, K., 2013. Location matters, especially for non-salient features–An eye-tracking study on the effects of web object placement on different types of websites. Int. J. Hum.-Comput. Stud. 71 (3), 228–235.

Saadé, R.G., Otrakji, C.A., 2007. First impressions last a lifetime: effect of interface type on disorientation and cognitive load. Comput. Hum. Behav. 23 (1) 525–535.

Salanitri, D., Hare, C., Borsci, S., Lawson, G., Sharples, S., Waterfield, B., 2015. Relationship between trust and usability in virtual environments: An ongoing study. In: Kurosu, M. (Ed.), Human-Computer Interaction: Design and Evaluation. Springer International Publishing, pp. 49–59.

Santini, M., 2008. Zero, single, or multi? Genre of web pages through the users' perspective. Inf. Process. Manage. 44 (2), 702–737.

Santini, M., Mehler, A., Sharoff, S., 2011. Riding the rough waves of genre on the web. In: Mehler, A., Sharoff, S., Santini, M. (Eds.), Genres on the Web: Computational Models and Empirical Studies. Springer Netherlands, Dordrecht, pp. 3–30. http://dx.doi.org/10.1007/978-90-481-9178-9_1.

Sauer, J., Sonderegger, A., 2011. The influence of product aesthetics and user state in usability testing. Behav. Inform. Technol. 30 (6), 787–796.

Seckler, M., Heinz, S., Forde, S., Tuch, A.N., Opwis, K., 2015. Trust and distrust on the web: User experiences and website characteristics. Comput. Hum. Behav. 45, 39–50.

Shepherd, M., Watters, C., 1998. The evolution of cybergenres. In: Proceedings of the Thirty-First Hawaii International Conference on System Sciences. Vol. 2. IEEE, pp. 97–109.

Skulmowski, A., Augustin, Y., Pradel, S., Nebel, S., Schneider, S., Rey, G.D., 2016. The negative impact of saturation on website trustworthiness and appeal: A temporal model of aesthetic website perception. Comput. Hum. Behav. 61, 386–393.

Sonderegger, A., Zbinden, G., Uebelbacher, A., Sauer, J., 2012. The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. Ergonomics 55 (7), 713–730.

Song, H., Schwarz, N., 2009. If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. Psychol. Sci. 20 (2), 135–138.

Strebe, R., 2016. Aesthetics on the web: effects on approach and avoidance behaviour. Behav. Inform. Technol. 35 (1), 4–20.

Thielsch, M.T., Blotenberg, I., Jaron, R., 2014. User evaluation of websites: From first impression to recommendation. Interact. Comput. 26 (1), 89–102.

Thielsch, M.T., Engel, R., Hirschfeld, G., 2015. Expected usability is not a valid indicator of experienced usability. PeerJ Comput. Sci. 1, e19.

Thielsch, M.T., Haines, R., Flacke, L., 2019. Experimental investigation on the effects of website aesthetics on user performance in different virtual tasks. PeerJ 7 e6516.

Tractinsky, N., 2013. Visual aesthetics. In: Soegaard, M., Dam, R.F. (Eds.), The Encyclopedia of Human-Computer Interaction, second ed. The Interaction Design Foundation, Aarhus.

Tractinsky, N., 2018. The usability construct: a dead end? Hum.–Comput. Interact. 33 (2), 131–177.

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., Sharfi, T., 2006. Evaluating the consistency of immediate aesthetic perceptions of web pages. Int. J. Hum.-Comput. Stud. 64 (11), 1071–1083.

Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. Interact. Comput. 13 (2), 127–145.

Tuch, A.N., Bargas-Avila, J.A., Opwis, K., 2010. Symmetry and aesthetics in website design: It's a man's business. Comput. Hum. Behav. 26 (6), 1831–1837.

Tuch, A.N., Presslaber, E.E., Stöcklin, M., Opwis, K., Bargas-Avila, J.A., 2012a. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. Int. J. Hum.-Comput. Stud. 70 (11), 794–811.

Tuch, A.N., Roth, S.P., Hornbæk, K., Opwis, K., Bargas-Avila, J.A., 2012b. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. Comput. Hum. Behav. 28 (5), 1596–1607.

Whitfield, T., 1983. Predicting preference for familiar, everyday objects: An experimental confrontation between two theories of aesthetic behaviour. J. Environ. Psychol. 3 (3), 221–237.

Whitfield, T.A., Slatter, P.E., 1979. The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. Br. J. Psychol. 70 (1), 65–75.

Willis, J., Todorov, A., 2006. First impressions: Making up your mind after a 100-ms exposure to a face. Psychol. Sci. 17 (7), 592–598.

Winkielman, P., Halberstadt, J., Fazendeiro, T., Catty, S., 2006. Prototypes are attractive because they are easy on the mind. Psychol. Sci. 17 (9), 799–806.

Ye, X., Peng, X., Wang, X., Teo, H.-H., 2020. Developing and testing a theoretical path model of web page impression formation and its consequence. Inf. Syst. Res. 31 (3), 929–949.

Aliaksei Miniukovich is a postdoctoral researcher at the Department of Information Systems, Production, and Logistics Management at the University of Innsbruck. He holds a Ph.D. in Information and Communication Technology from the University of Trento. Aliaksei's research focuses on the user's first impression and preferences and often relies on quantitative methods and computational models. He also serves as an associate editor for the International Journal of Human-Computer Studies and frequently reviews for a variety of HCI-related journals and conferences.

Kathrin Figl is an Associate Professor of Information Systems at the University of Innsbruck. She holds a venia legendi for Information Systems from the Vienna University of Economics, where she served as an Assistant Professor at the Institute for Information Systems and New Media. She has earned two Master's Degrees in Information Systems and Psychology, as well as a Doctoral Degree in Information Systems, all with honors, from the University of Vienna. Her research focuses on human-centric development and design of information systems, and she has published more than 80 research papers and articles, for example, in the Journal of Management Information Systems, the Journal of the Association for Information Systems, Decision Support Systems, and Information & Management. She served as track chair for the track "Cognition and Human Behavior in Information Systems" at ECIS 2019–2023.