The aesthetic pleasure in design scale

The development of a scale to measure aesthetic pleasure for designed artifacts

Blijlevens, Janneke; Hekkert, Paul; Thurgood, Clementine; Chen, Lin-Lin; Allen Whitfield, T.W.

DOI
10.1037/aca0000098

Publication date
2017

Document Version
Final published version

Published in
Psychology of Aesthetics, Creativity, and the Arts

Psychology of Aesthetics, Creativity, and the Arts
2017, Vol. 11, No. 1, 86–98

1931-3896/17/$12.00 http://dx.doi.org/10.1037/aca0000098

The Aesthetic Pleasure in Design Scale: The Development of a Scale to Measure Aesthetic Pleasure for Designed Artifacts

Janneke Blijlevens
Royal Melbourne Institute of Technology University

Clementine Thurgood
University of Technology Sydney

Paul Hekkert
Delft University of Technology

Lin-Lin Chen
Eindhoven University of Technology

Helmut Leder
University of Vienna

T. W. Allan Whitfield
Swinburne University of Technology

There is a lack of consistency regarding the scales used to measure aesthetic pleasure within design. They are often chosen ad hoc or adopted from other research fields without being validated for designed artifacts. Moreover, many scales do not measure aesthetic pleasure in isolation, but instead include its determinants (e.g., novelty). Therefore, we developed a new scale to measure aesthetic pleasure and included scales to measure its known determinants for discriminant validity purposes, which automatically led to validating these determinants as well. In the exploratory phase, we identified highly reliable items representative of aesthetic pleasure and its determinants across product categories. In the validation phase, we confirmed these findings across different countries (Australia, the Netherlands, and Taiwan). The final scale consists of 5 items, "beautiful," "attractive," "pleasing to see," "nice to see," and "like to look at," that together reliably capture the construct of aesthetic pleasure. Several recommendations are formulated regarding the application of this scale in design studies and beyond.

Keywords: aesthetic pleasure, design, scale development, determinants of aesthetic pleasure

Research into aesthetic pleasure or appreciation is often confined to art perception and appreciation (Hekkert, 2014b). Although works of art are—or should we say "were"—often created to delight the perceiver, for beauty purposes, they are clearly not the only "objects" that can be pleasant to look at, listen to, or touch. We can aesthetically appreciate a landscape or a photograph of that same landscape; we find beauty in faces, buildings, and other man-made things; we can even be aesthetically pleased by, and therefore ascribe beauty to, an idea, a chess move, or a scientific proof (Da Silva, Crilly, & Hekkert, 2015). Any object can be aesthetically appreciated, and objects are often deliberately designed to induce aesthetic pleasure (Postrel, 2003). Accordingly, we see an increasing interest in researching aesthetic pleasure derived from everyday objects such as products and websites in design research, consumer research, and human-computer interaction (HCI) research (e.g., Blijlevens, Carbon, Mugge, & Schoormans, 2012; Hassenzahl & Monk, 2010; Hekkert, Snelders, & Van Wieringen, 2003).

While ample research into what people find aesthetically pleasing exists in design, marketing, arts, and psychology literature (e.g., Blijlevens et al., 2012; Bloch, 1995; Hekkert, 2006, 2014a, 2014b; Hekkert et al., 2003; Hoyer & Stokburger-Sauer, 2012; Leder, Belke, Oeberst, & Augustin, 2004; Leder, Ring, & Dressler, 2013; Schoormans & Robben, 1997; Swami, 2013; Veryzer & Hutchinson, 1998) research into how aesthetic pleasure for designed artifacts should actually be defined and subsequently be measured has received little attention. More specifically, in the design field, most research focuses on how certain determinants

Janneke Blijlevens, Behavioural Business Lab, School of Economics, Finance and Marketing, Royal Melbourne Institute of Technology University; Clementine Thurgood, Design Innovation Research Centre, University of Technology Sydney; Paul Hekkert, Department of Industrial Design Engineering, Delft University of Technology; Lin-Lin Chen, Department of Industrial Design, Eindhoven University of Technology; Helmut Leder, Faculty of Psychology, University of Vienna; T. W. Allan Whitfield, Centre for Design Innovation, Swinburne University of Technology.

Project Unified Model of Aesthetics is supported by Maatschappij- en Gedragswetenschappen VICI Grant 453-10-004 from the Netherlands Organization for Scientific Research (NWO), awarded to Paul Hekkert. We would like to express our gratitude to Wei-Ken Hung and Ya-Han Lee for translating the questionnaire into Mandarin and deploying that questionnaire to the Taiwanese sample, Ruben Post for his help with item generation, Shivani Tyagi for creating the experimental stimuli, and Marc Hassenzahl and Nathan Crilly for their continuing conceptual, theoretical, and practical input throughout this project.

Correspondence concerning this article should be addressed to Janneke Blijlevens, Behavioural Business Lab, School of Economics, Finance and Marketing, Royal Melbourne Institute of Technology University, Building 80, 445 Swanston Street, Melbourne, 3000, Victoria, Australia. E-mail: janneke.blijlevens@rmit.edu.au

such as typicality, novelty, complexity, unity, and variety explain variation in evaluations of aesthetic pleasure (e.g., Blijlevens et al., 2012; Hekkert et al., 2003; Moshagen & Thielsch, 2010; Whitfield & Slatter, 1979). Results of these studies are, however, hard to compare and aggregate because of the different ways in which the dependent variable aesthetic pleasure has been operationalized. Scales are often chosen ad hoc or do not measure aesthetic pleasure as such, but instead include its determinants (e.g., typicality, symmetry) as constituents of aesthetic pleasure (see, e.g., Augustin, Carbon, & Wagemans, 2011; Hassenzahl, Burmester, & Koller, 2003). These scales can indicate whether a given object is expected to be generally pleasing, but we lack a reliable and valid scale to actually measure aesthetic pleasure as distinct from its determinants. In order to accurately establish which factors influence aesthetic pleasure, and how it is that they exert their influence it is essential to measure and treat aesthetic pleasure and its determinants separately. Because of the noted diversity in scales used in the literature, and consequent noncomparability of findings, a systematic bottom-up approach is required along three structured scale validation phases. First, we perform an in-depth literature analyses and gain expert advice to identify suitable and relevant items to measure aesthetic pleasure for designed artifacts. Second, we follow with two phases of systematic scale optimization and delimitation of related constructs.

Measures of Aesthetic Pleasure

Within the area of aesthetics research into designed artifacts, many of the scales used to measure aesthetic pleasure are chosen ad hoc or are chosen based on previous studies of aesthetic pleasure, and were not empirically tested to determine whether they do reliably and validly measure aesthetic pleasure. For example, many researchers refer back to Page and Herr who used "attractive" as an item to measure aesthetic pleasure (Page & Herr, 2002). Others opt for items such as "beautiful," "pleasing," and "liking" (e.g., Hung & Chen, 2012; Martindale, Moore, & Borkum, 1990). In those cases, often no sources from which the items were taken are included so it is unclear whether they came from validated scales, nor is it clear whether they were derived from a comprehensive theoretical approach. Even if these items are appropriate measures of aesthetic pleasure, they have not been tested for reliability and validity, making comparisons between studies in design aesthetics difficult.

In HCI, several scales have been developed to measure appreciation of websites and interactive products. For example, the scale AttrakDiff is now widely used (Hassenzahl & Monk, 2010). This scale measures "pragmatic value," "hedonic value," "beauty," and "goodness." In particular, hedonic value is described to assess aesthetic pleasure. Items that measure hedonic value include "captivating," "stylish," "premium," and "creative." Also within HCI, scales were developed that specifically focus on aesthetic pleasure for web designs using items such as "the layout is too dense," "the colors are attractive," "the layout is pleasantly varied" (Moshagen & Thielsch, 2010), and "pleasing," "sophisticated," "symmetrical," and "modern" (Lavie & Tractinsky, 2004). In the field of art, a scale is being developed that aims to measure aesthetic pleasure for artworks, and this includes items such as "beautiful," "incomprehensible," "fascinating," "ordinary," "original," "innovative," "attractive," "happy," "warm," and "overwhelming" (Augustin et al., 2011).

A significant shortfall of these existing scales in HCI and art is that these scales include items that generally measure determinants of aesthetic pleasure but do not measure aesthetic pleasure "as such"; that is, as a singular, separately defined construct. Items such as "innovative," "original," and "ordinary," for example, are used in other studies to measure novelty and typicality; factors shown to be important predictors of aesthetic pleasure (e.g., Hekkert et al., 2003). A large body of research in design aesthetics investigates which design factors (e.g., novelty) increase aesthetic pleasure (e.g., Blijlevens et al., 2012; Hekkert, 2006, 2014a, 2014b; Veryzer & Hutchinson, 1998; Whitfield & Slatter, 1979). This type of research provides insights into the psychological and cognitive mechanisms underlying the aesthetic pleasure for products, as well as practical implications for designers and marketers; however, in order to substantiate the claims made regarding the relationships of design factors with aesthetic pleasure, these factors need to be measured separately from aesthetic pleasure. This is very clearly illustrated with an item used in Moshagen and Thielsch (2010): "the layout is pleasantly varied" wherein both pleasant, which could measure aesthetic pleasure, and varied, known to be a determinant for aesthetic pleasure, are combined into one item. If the goal is to assess how variety in a design and aesthetic pleasure for that design are related then their validated scale cannot be employed. Therefore, we set out to develop a scale that not only measures aesthetic pleasure in isolation, but also separates this construct clearly from its determinants. Before being able to measure the concept "aesthetic pleasure" adequately, we, therefore, need to define it unambiguously.

Defining Aesthetic Pleasure

If one aims to develop a scale to measure a psychological construct it is crucial to first define it as precisely as possible. We adopt the following definition of aesthetic pleasure: "the pleasure people derive from processing the object *for its own sake*, as a source of immediate experiential pleasure in itself, and not essentially for its utility in producing something else that is either useful or pleasurable" (Dutton, 2009, p. 52). Following this definition, people can find it aesthetically pleasing to watch a sunset or feel the curves of a Ferrari, people can find beauty in the latest Koolhaas building or derive aesthetic pleasure from listening to a classic Beatles song; in fact, people can even aesthetically appreciate the most mundane things, such as the graphic layout on a package of cigarettes. More recently, Hekkert (2014a, p. 278) argued along similar lines that this aesthetic pleasure ". . . is limited to the gratification that comes from sensory perception of an object or any other stimulus, including abstract ideas. . . ." The aesthetic pleasure we refer to here is therefore not limited to visual gratification, but applies to all sensory domains (Hekkert & Schifferstein, 2008). Furthermore, aesthetic pleasure "has no direct implications for any of our everyday concerns, the class of dispositional states that *is* fundamental to our emotions" (Hekkert, 2014a, p. 278). According to the most dominant theory in emotion psychology, appraisal theory, an emotion is elicited by an appraisal of an event or situation as potentially beneficial or harmful to a person's concerns (e.g., Scherer, Schorr, & Johnstone, 2001). By contrast and in line with Dutton's definition, it has repeatedly been

argued that an aesthetic response is "disinterested" (Kant, 1790/1952) or distanced (Bullough, 1912) in that no motive other than perceiving the object of perception, as such, is involved. This is not to say that recognizing an object's purpose cannot induce aesthetic pleasure; rather it says that actual fulfillment of a need or actual use of the object is not a prerequisite for an aesthetic response (Hekkert, 2014b). For those reasons, and strictly speaking, an aesthetic experience is not an emotion (Hekkert, 2014a).

The definition of aesthetic pleasure adopted here is thus a narrower one than the "aesthetic response" used in art studies, which can refer to an array of emotional and cognitive experiences that people have when perceiving a sculpture or painting. In this context, some also speak of "aesthetic emotions": the range of emotions, such as awe, fascination, bewilderment, sadness, and so on, that people may go through when processing a work of art (e.g., Frijda, 1989). Take, for instance, the often cited aesthetic process model of Leder et al. (2004). This model describes how people process a work of art and what the outcomes of this processing can be. The complete combination of cognitive and affective processes leads to a result in the form of an aesthetic episode, response, and judgment, such as "this is an interesting painting" or "this painting moves me." In the context of art, aesthetic pleasure is only one facet of the full aesthetic response as documented by Leder and his colleagues, and many others (e.g., Cupchik & Laszlo, 1992; Kreitler & Kreitler, 1972; Leder & Nadal, 2014). Paintings are often deliberately created to elicit an evocative aesthetic response, while for designed artifacts aesthetic pleasure is often the *only* aesthetic response people have, next to experiences related to, for instance, affordances, usability, and expressive meaning (Hekkert & Schifferstein, 2008; Norman, 1988). This does not mean that aesthetic responses are of minor importance in the design context. To the contrary, attractive products appear, for example, more usable and to be of increased value (see Hekkert, 2014a for an overview; Bloch, 1995; Tractinsky, Katz, & Ikar, 2000). The field of product design thus demands and conveniently allows for studying aesthetic pleasure in the "pure" sense as we have defined it.

The Current Research

In this research, we set out to create a validated, reliable, and generalizable scale to measure aesthetic pleasure for designed artifacts. We report three research phases: an item generation phase (with items being the different questions used to measure a construct using Likert scales: e.g., "this is attractive" is an item measuring aesthetic pleasure), an exploratory phase, and a validation phase. In the item generation phase, we collected items to measure aesthetic pleasure, assessing their relevance for design, and rewording them into Likert scale-type items pertaining to designed artifacts. In the exploratory phase, we investigated how the different items load on our intended constructs through an exploratory factor analysis by analyzing the data from respondents rating different product designs using the items that were identified to measure aesthetic pleasure in the item generation phase. We then assessed the complete structure as well as all constructs separately. In addition, comparisons of factor structures between product categories and a retest reliability study were performed. Factor model validation was performed in the validation phase. A confirmatory factor analysis was performed using structural equation modeling (SEM), wherein the resulting factor model from the exploratory phase was then tested on new samples of respondents taken from three different countries (Australia, the Netherlands, and Taiwan) and included stimuli from two new sets of product categories than those used in the exploratory phase. The research used stimuli from several different product categories and, within those product categories, several different designs were presented that together represent the wide variety of designs possible within that product category. That way, we aim to assure generalizability of our scale across designed artifacts. In addition, to assess convergent validity and discriminant validity in the exploratory and validation phases, next to items measuring aesthetic pleasure, items intended to measure its determinants (typicality, novelty, unity, and variety) were also included. Because these constructs are assumed to be indicators of aesthetic pleasure, we expected them to be separate factors from aesthetic pleasure in a factor solution, and to positively affect aesthetic pleasure in a path model. A beneficial consequence of this procedure meant that we were also able to validate scales for these determinants. Finally, in the validation phase, discriminant validity with product emotion (Desmet, 2003) and product usability (adapted from Spangenberg, Voss, & Crowley, 1997) was also assessed.

Item Generation Phase

First Phase

Three experts in design research performed an extensive review of English written literature discussing, theorizing, and empirically investigating aesthetic pleasure in the fields of design, arts, HCI, perception psychology, and consumer psychology. All researchers made lists of items measuring the construct as used in the literature. These items were collected and carefully studied to remove replicates. This left 86 items that were individually transferred onto Post-its for further processing (see all items in Appendix A).

Second Phase

Two researchers familiar with the literature on design aesthetics categorized all the items that were written on individual Post-its into two categories: "aesthetic pleasure" or "determinant of aesthetic pleasure." In making these decisions, the researchers considered whether the items adequately reflected our construct of interest, aesthetic pleasure (as defined in the preceding section, Defining Aesthetic Pleasure) as a specific response, or whether they reflected constructs known to influence aesthetic pleasure. This categorization process resulted in 37 items for aesthetic pleasure and 49 items that were considered determinants of aesthetic pleasure (as displayed in Appendix B). Examples of determinants include "familiar," "novel," "understandable," "patchy," and "fluent to process." The 37 items for aesthetic pleasure were then used as input for a second categorization task wherein the researchers rated the items on their relevance to the concept aesthetic pleasure on a scale from 1 to 5 (1 = *not at all relevant*, 5 = *very relevant*). When the researchers did not agree, they discussed each item until they reached a communal decision. The items that received a score of three or above were then used as input for the third phase of item generation (23 items—see Appendix C).

Third Phase

Seven established researchers in aesthetics with different specializations (i.e., design, HCI, psychology, and the arts) rated all 23 items on the level to which they thought these items were representative of the construct aesthetic pleasure by using a web-based questionnaire. Researchers from different disciplines were approached to ensure the final items would be generalizable to all kinds of manmade artifacts. However, to make sure that all respondents had the same goal of research in mind they received the following instruction before rating the items (on 5-point Likert scales):

> In order to decide on what items measure the construct 'aesthetic pleasure' we are going to ask you to rate several items on how well they measure aesthetic pleasure (relevance, practicality, content-correct) according to you. When rating the items it is important to keep a few things in mind:
>
> 1) We are aware of the fact that there exist questionnaires that measure aesthetic pleasure. These questionnaires include measures such as pleasurable and likable, but also items such as novel, dynamic, unified and complex. We believe that questionnaires that include items such as novel, dynamic, unified and complex have a good predictive value of whether a product or an interface will be aesthetically pleasing. However, we are not necessarily interested in predicting whether a product will be aesthetically pleasing, but we want to know how such factors as novelty, dynamic, unified and complex influence aesthetic pleasure. In order to be able to perform research that provides such insights, we have to separate measures for novelty and so forth from measures of aesthetic pleasure. Hence, factors that influence aesthetic pleasure (antecedents) are not included in this current questionnaire that we send you.
>
> 2) The items should be able to measure aesthetic pleasure for objects on all sensory domains: touch, sound, vision, taste. Please, keep this in mind when rating the items.
>
> Next you will see all the items and we ask you to rate them on the level in which you think these items are a good measure (relevance, practicality, content-correct) of aesthetic pleasure on a scale from 1–5.

The average scores of each item (see Appendix C) were then used as qualitative input for an extensive discussion between five of the researchers. They were instructed to pay particular attention to whether the items were relevant to the construct of aesthetic pleasure, and whether they were also sufficiently different to each other in conceptual meaning so that the full construct of aesthetic pleasure could be captured (Rossiter, 2002). This resulted in the five final items: pleasant, attractive, nice, beautiful, and like. These items were then reworded with the help of two researchers with English as their first language to ensure relevance for measuring aesthetic pleasure for designed artifacts. The final items used as input for the exploratory study were: ". . . this is a beautiful [object (e.g., camera)]," ". . . this is an attractive [object]," ". . . this [object] is pleasing to see," ". . . this [object] is nice to see," and ". . . I like to look at this [object]."

A similar item generation procedure was performed for items measuring the constructs typicality, novelty, unity, and variety. The final items for these determinants were ". . . this is a typical [object (e.g., camera)]," ". . . this is representative of a [object]," ". . . this design is common for a [object]," ". . . this is a standard design," and ". . . this is characteristic of a [object]" for typicality; ". . . this is a novel [object]," ". . . this design is original," ". . . this is a new example of a [object]," and ". . . this design is innovative" for novelty; "this is a unified design," "this is a coherent design," and "this is an orderly design" for unity, and "this design is rich in elements," "this design is made of different parts," and "this design conveys variety" for variety.

Exploratory Phase

Method

Stimuli selection. A total of 20 stimuli (Product Category × Product Design) were rated by our respondents. Images from four different product categories were chosen as stimuli (cameras, motorcycles, chairs, and websites) to ensure that aesthetic pleasure was generalizable across a broad range of product categories. To ensure robustness of our results, within each product category five designs were selected to represent the variety of potential designs found within that product category. Images were edited where necessary so that any identifying brand features and text were removed.

Respondents. A total of 157 respondents from Australia participated in this research. Respondents were recruited from a consumer panel instead of a student population for generalizability purposes. Respondents received reward points for participation that can be exchanged for goods in an online shop when enough reward points are saved; a common compensation for respondents from this consumer panel. Of these 157 respondents, answers were not considered from people who did not finish the questionnaire and who did not have English as their first language. Finally, the respondents' answers were checked and all respondents that only answered extreme values (1 or 7), only neutrals (4) or only consecutive responses (e.g., 2,2,2 . . . 2,2,2) were deleted from the analyses. The final analyses were performed with a total of 108 respondents (mean age = 52, SD = 13, 66 females).

Procedure. Respondents were informed that they would be asked to view and rate a series of images of products. Upon presentation of each image, they were asked to indicate the extent to which they agreed with a series of statements describing each given design using 7-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). The aforementioned final items from the generation phase were used for aesthetic pleasure and the items representing its commonly investigated determinants—typicality, novelty, unity, and variety—were used to assess the discriminant validity of the aesthetic pleasure scale. Product designs and order of rating scales were presented in random order, at a participant-paced interval using a web-based questionnaire.

Results

All data analyses were performed with a nonaggregated dataset. Intraclass correlations (ICCs) between the aesthetic pleasure ratings were very low, <.20 (ICC = .084), which is why we can conclude that people did not agree on the level to which they rated designs, even though significance was achieved ($p < .001$). Therefore aggregation would diminish a lot of the unique information present in the dataset.

Exploratory factor analysis. An exploratory factor analysis with Varimax rotation revealed five separate factors (based on

eigenvalues >1.0): aesthetic pleasure, typicality, novelty, unity, and variety. Two items were deleted from the final structure because they did not conceptually fit with the factor they loaded highest upon: "good example of the category" (conceptually belonging to the construct typicality) and "diverse" (conceptually belonging to the construct variety).

Reliability. Cluster analysis revealed that all correlations were above .50 and significant, so all items were retained. Factor invariance analysis showed no significant differences between product categories for each factor. Cronbach's alphas were .98 for aesthetic pleasure, .87 for novelty, .93 for typicality, .90 for unity, and .83 for variety.

Retest reliability. To assess retest reliability, a subsample of the previous sample ($N = 50$) was administered the exact same questionnaire again after a week's time had passed. All correlations between Time 1 and 2 for each item were above .5 and significant, except for the item "different parts" loading on the construct variety (.463). Given that this item loaded the highest on the factor variety, to which it conceptually belongs, and did not show significant differences across the product categories (the invariance analysis revealed that it loaded highest on the factor variety for all product categories), we decided not to exclude it. All correlations between the factors at Time 1 and Time 2 were significant and mostly higher than the recommended level of .7 (Nunnally, 1978), except for unity (.659) and variety (.584). Given that the remaining correlations were very high, particularly for our construct of interest—aesthetic pleasure, we decided that retest reliability was sufficient to enter all five factors and their items into the factor model tested in the next validation phase of this research.

Validation Phase

Method

Stimuli selection. A total of 20 stimuli (Product Category × Product Design) were rated by our respondents. For replication purposes, two product categories used in the exploratory phase were used as stimuli in the validation phase: cameras and chairs. For generalization purposes two new product categories were added: sunglasses and sanders. We chose these two additional product categories because we wanted to be able to validate our results from the exploratory phase using product categories that differ in symbolic, functional, and ergonomic value (Creusen & Schoormans, 2005). As in the previous phase, within each product category five designs were selected to represent the wide variety of designs found within that product category.

Respondents. Respondents from consumer panels from three different countries (Australia, the Netherlands, and Taiwan) participated in this research. As before, respondents' answers were not considered in the analyses for people who did not finish the questionnaire and who did not have English (for the Australian sample), Dutch (for the Dutch sample), or Mandarin (for the Taiwanese sample) as their first language (see Appendix D for all items in all three languages). Finally, the respondents' answers were checked and all respondents that only answered extreme values, only neutrals, or only consecutive responses were deleted from the analyses. The final analyses were performed with a total of 591 participants (200 from Australia, mean age = 46, $SD = 16$, 113 females; 200 from the Netherlands, mean age = 50, $SD = 14$, 131 females; and 191 from Taiwan, mean age = 21, $SD = 15$, 129 females).

Procedure. Again, respondents viewed and rated a series of images of products using a web-based questionnaire. Upon presentation of each image, they were asked to indicate how much they agreed with statements describing the given designs using 7-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). In this phase, the items that served as final output from the exploratory phase were used in the validation phase.

Since we conducted this validation phase in three different countries (Australia, the Netherlands, and Taiwan) the items had to be translated into the different languages. Four Dutch and four Taiwanese researchers participated in the translate-back-translate process. For each country, two researchers were involved in the project, thereby assuring face validity of the construct being measured and the other two were independent ensuring language objectivity and avoiding use of jargon associated with the field of aesthetics. First, one involved and one independent researcher for each country translated the English items into their respective language (Dutch or Mandarin) and discussed the items until they agreed on the best translation. Then the translated items were back-translated into English by the two other (one independent and one involved for each country) researchers, without knowing what the English items were. The researchers were then presented with the original English items and where there was a mismatch or disagreement in the back-translations, the researchers discussed until they agreed upon the best Dutch/Mandarin translation.

A balanced design was used wherein respondents were randomly assigned to start with one of the four product categories, in which each design and their ratings scales were randomly presented. In addition, respondents rated the product designs on semantic descriptions taken from the 14-item Product Emotion scale (PrEmo; Desmet, 2003) and on items measuring usability (taken from Spangenberg et al., 1997): "this [object] seems useful," "this design seems practical," "this [object] seems functional," "this design seems sensible," and "this [object] seems handy." The PrEmo items were taken directly from Desmet (2003), but differed in the sense that they were verbal descriptions, and not animated pictures, for consistency within the current research (i.e., the Likert scale format). This decision was acceptable because the PrEmo scale initially comprised descriptive items and pictorials were only added after these initial items were proven to be effective in measuring product emotions (Desmet, 2003).

Results

All data analyses were performed with a correlation matrix used as input in AMOS 22 (Arbuckle, 1995) for SEM.

Confirmatory factor analysis. SEM was used to assess whether the input model that resulted from the exploratory phase was structurally confirmed with the results of the sample from the validation phase. In other words, the same factors should underlie the items of the second sample as in the input model based on the sample of the exploratory phase. The five-factor model (aesthetic pleasure, typicality, novelty, unity, and variety) from the exploratory phase was used to test the data obtained in the second study by means of the two-step approach of SEM described by Anderson and Gerbing (1988).

The output file generated through SEM executed by AMOS provided fit measures and suggested no modifications to the model were needed. The results validated the five-factor model that resulted from the exploratory phase: the goodness of fit measure (GFI) was 0.917; the normed fit index (NFI) was 0.953; the comparative fit index (CFI) was 0.954, and the adjusted goodness of fit measure (AGFI) was .891. Additionally, the root-mean-square error of approximation (RMSEA) showed an acceptable fit (0.07; acceptable: 0.05 < RMSEA > 0.08; Jais, 2006). All items had statistically significant loadings on their factors and varied between 0.60 and 0.95, which is consistent with the five-factor model taken as input from the exploratory phase of the research. All explained variances (squared multiple correlations [SMCs]) of our items varied between .40 and .86. The final five-factor model is depicted in Table 1.

Reliability and convergent validity. The average variance extracted (AVE) for each attribute was higher than 0.50, which indicates convergent validity (Table 2). Composite reliability of the attributes was assessed with the Fornell and Larcker (1981) criterion. All attribute reliability measures were high (lowest was 0.79 for variety; see Table 2).

Discriminant validity within the model. The model's discriminant validity between the five constructs was deemed to be good because a chi-square test between the model in which the construct correlations were constrained to be 1.0 and the unconstrained model proved to be significant (Jöreskog, 1971). This means that constraining the model to 1.0 made the fit for the model significantly worse. Moreover, all squared interconstruct correlations (Table 3) were lower than the AVEs, which indicates discriminant validity between the constructs; that is, each construct has its own explained variance separate from the other constructs.

Nomological validity. As expected, all interconstruct correlations between aesthetic pleasure and its determinants were positive and significant (all >.36; Table 4).

Table 2
The Average Variance Extracted and Fornell and Larcker Reliability Criterion for Aesthetic Pleasure and Its Determinants Typicality, Novelty, Unity, and Variety

Factors	AVE	FR
Aesthetic pleasure	.84	.96
Typicality	.69	.92
Novelty	.68	.89
Unity	.68	.87
Variety	.57	.79

Note. AVE = average variance extracted; FR = Fornell and Larcker.

Determinants' predictive ability of aesthetic pleasure. In the literature, the predictive relationships of determinants like typicality, novelty, unity, and variety with aesthetic pleasure are often the focus of research. Therefore, we deemed it important to assess whether these determinants were indeed predictive of aesthetic pleasure. A model was tested wherein paths were drawn between all determinants and aesthetic pleasure to assess whether, as can be theorized, all determinants significantly influence aesthetic pleasure. The model showed a good fit (chi/df = 54.420, $p <$ .001, GFI = .917, NFI = .953, CFI = .954, AGFI = .891, RMSEA = .070).

All regression weights were significant and positive. The standardized regression weights were higher than .6 for all items with their relevant construct. The standardized regression weights for aesthetic pleasure with the determinants typicality, novelty, unity, and variety were .247, .211, .438, and .237, respectively. Hence, we can assume predictive value of our determinants with aesthetic pleasure.

Group comparison between countries. Group comparisons between countries (Taiwan, the Netherlands, and Australia)

Table 1
The Standardized Regression Weights for Aesthetic Pleasure, Typicality, Novelty, Unity, and Variety in Bold Type Face

Items	Aesthetic pleasure	Typicality	Novelty	Unity	Variety
Like to look	**.919**	.000	.000	.000	.000
Nice to see	**.927**	.000	.000	.000	.000
Pleasing to see	**.924**	.000	.000	.000	.000
Attractive	**.912**	.000	.000	.000	.000
Beautiful	**.904**	.000	.000	.000	.000
Standard	.000	**.873**	.000	.000	.000
Common	.000	**.873**	.000	.000	.000
Representative	.000	**.730**	.000	.000	.000
Typical	.000	**.874**	.000	.000	.000
Characteristic	.000	**.776**	.000	.000	.000
Innovative	.000	.000	**.844**	.000	.000
New example	.000	.000	**.849**	.000	.000
Original	.000	.000	**.783**	.000	.000
Novel	.000	.000	**.817**	.000	.000
Coherent	.000	.000	.000	**.810**	.000
Orderly	.000	.000	.000	**.860**	.000
Unified	.000	.000	.000	**.804**	.000
Conveys variety	.000	.000	.000	.000	**.810**
Different parts	.000	.000	.000	.000	**.631**
Rich in elements	.000	.000	.000	.000	**.799**

Table 3
The Squared Interconstruct Correlations for Aesthetic Pleasure and Its Determinants Typicality, Novelty, Unity, and Variety

Factors	AVE	Squared interconstruct correlations (standardized model)			
		Aesthetic pleasure	Typicality	Novelty	Unity
Aesthetic pleasure	.84				
Typicality	.69	.28			
Novelty	.68	.13	.13		
Unity	.68	.56	.49	.023	
Variety	.57	.36	.01	.51	.19

Note. AVE = average variance extracted.

showed that the five-factor model that was found in the exploratory phase of the research and was confirmed in the validation phase, fits for the Taiwanese, Dutch, and Australian samples (chi/df = 25.102, $p < .001$, GFI = .886, NFI = .938, CFI = .940, AGFI = .856, RMSEA = .047). For the Taiwanese sample, all regression weights were significant and >.7 except for the regression weight of "characteristic" predicting the construct typicality (.539), all correlations were significant and varied between .160 and .767, and all SMCs varied between .291 and .817. For the Dutch sample, all regression weights were significant and above .7, all correlations were significant and varied between .163 and .754, and all SMCs varied between .536 and .895. For the Australian sample, all regression weights were significant and >.6 except for the regression weight of "different parts" predicting the construct variety (.575), all correlations were significant and varied between .163 and .689, and the SMCs varied between .331 and .893.

A chi-square difference test showed that the model in which equal regression weights between groups were assumed had a significantly worse model fit than when regression weights were allowed to differ between countries. This means that even though the items can be used to measure the five intended constructs for each country, there are differences in how much some items contribute to a certain construct between countries. This can be due to translation issues, but also due to how common certain words are in the language itself.

Discriminant validity with product emotions. Aesthetic pleasure has an inherent positive connotation (Desmet & Hekkert, 2007); hence we expected a positive relationship between aesthetic pleasure and positive emotions, but a negative relationship between aesthetic pleasure and negative emotions. First, an exploratory factor analysis was performed on the 14 product emotions (PrEmo; Desmet, 2003). Based on eigenvalues and scree-plot analysis two factors were extracted: positive valence and negative valence. This is congruent with the circumplex model of emotion in which valence is considered the first and main dimension on which emotions differ (Russell, 1980; Russell & Barrett, 1999). The Cronbach's alphas for these factors were: .934 for positive valence and .917 for negative valence.

In AMOS, aesthetic pleasure, positive valence, and negative valence were included in a model to assess discriminant validity. The total model showed a good fit (chi/df = 53.877, $p < .001$, GFI = .914, NFI = .958, CFI = .959, AGFI = .890, RMSEA = .070). All regression weights were significant and all standardized regression weights >.6 for all items on each construct.

Intercorrelations were significant and in the expected directions. Aesthetic pleasure and positive valence had a positive correlation of .72, aesthetic pleasure and negative valence had a negative correlation of −.47, and positive and negative valences had a correlation of −.12.

The model's discriminant validity (between the three constructs: aesthetic pleasure, positive valence, negative valence) was found to be good because a chi-square test between the model in which the construct correlations were constrained to be 1.0 and the unconstrained model proved to be significant (Jöreskog, 1971). This means that constraining the variances to 1 made the fit for the model significantly worse.

All squared interconstruct correlations (see Table 5) are lower than the AVEs (for aesthetic pleasure: .84, positive emotions: .68; and negative emotions: .62), which indicates discriminant validity between the constructs (each construct has its own explained variance separate from the other constructs).

Discriminant validity with product usability. Aesthetically pleasing products are often also easier to understand and therefore considered useful or usable (Hekkert, 2014a). Usability and aesthetic pleasure are thus related, but are two separate factors in which the

Table 4
Intercorrelations Between Aesthetic Pleasure, Typicality, Novelty, Unity, and Variety

Factors	Aesthetic pleasure	Typicality	Novelty	Unity
Aesthetic pleasure				
Typicality	.53			
Novelty	.36	−.36		
Unity	.75	.70	.15	
Variety	.60	.10	.71	.44

Table 5
Squared Interconstruct Correlations (Standardized Model) and AVEs for Aesthetic Pleasure, Positive Emotion, and Negative Emotion

Factors	AVE	Aesthetic pleasure	Positive emotion
Aesthetic pleasure	.84		
Positive emotion	.68	.51	
Negative emotion	.62	.22	.02

Note. AVE = average variance extracted.

underlying items of each should measure two separate constructs. We assessed discriminant validity to assess whether our measure of aesthetic pleasure is indeed a separate construct from product usability. However, based on previous research, we expected a positive relationship between aesthetic pleasure and product usability.

First, an exploratory factor analysis was performed for product usability. Based on eigenvalues and scree plot, one factor was extracted. The Cronbach's alpha was .958.

In AMOS, aesthetic pleasure and product usability were included in one model to assess discriminant validity. The total model showed a good fit (chi/df = 33.221, $p < .001$, GFI = .979, NFI = .991, CFI = .992, AGFI = .967, RMSEA = .055). All regression weights were significant and the standardized regression weights were >.8 for all items on each construct. Intercorrelations were significant and in the expected direction: aesthetic pleasure and product usability had a correlation of .733.

The model's discriminant validity (between the two constructs: aesthetic pleasure and product usability) is good because a chi-square test between the model in which the construct correlations were constrained to be 1.0 and the unconstrained model proved to be significant (Jöreskog, 1971). This means that constraining the correlations to 1 made the fit for the model significantly worse.

The squared interconstruct correlation (r^2 = .54) is lower than the AVEs (for aesthetic pleasure: .84, and for product usability: .82), which indicates discriminant validity between the constructs (each construct has its own explained variance separate from the other construct).

General Discussion

In the introduction it was argued that research within the domain of design aesthetics lacks a valid scale to measure the construct of interest: aesthetic pleasure. Thus, this research set out to develop a reliable, valid, and generalizable scale to measure aesthetic pleasure in the domain of design. We found that aesthetic pleasure can be validly and reliably measured with five items: ". . . this is a beautiful [object (e.g., camera)]," ". . . this is an attractive [object]," ". . . this [object] is pleasing to see," ". . . this [object] is nice to see," and ". . . I like to look at this [object]." These items measure the construct aesthetic pleasure and clearly separate it from its determinants. Thus, this scale can be used in further empirical studies in design aesthetics that aim to assess the factors determining aesthetic pleasure. The scale was also deemed valid and reliable for different countries, including both Western and Eastern countries. Furthermore, we defined aesthetic pleasure as a direct response to an object, which often precedes judgments of its utilitarian qualities or the needs it can fulfill. Indeed, we managed to capture and measure the aesthetic response to designs as separate from an emotional or cognitive response, as indicated through discriminant validity with the product emotion (adapted from Desmet, 2003) and usability scales (adapted from Spangenberg et al., 1997). Hence, we created a scale that measures the immediate pleasurable response people have toward designed objects in their environment, as distinct from other types of more considered responses.

As a consequence of this scale validation study, we also identified items suitable to measure some prominent determinants of aesthetic pleasure: typicality, novelty, unity, and variety (see Appendix D). These items were tested for reliability and validity and were also deemed generalizable across cultures and product categories. Identification of these items opens up possibilities to reliably assess their (combined) effects on aesthetic pleasure. Consequently, in future studies the seemingly controversial effects of these determinants on aesthetic pleasure can be resolved.

The final scale may not come as a surprise to some as several of the items identified and validated to measure aesthetic pleasure in this research (e.g., "beautiful" and "attractive") are the same as the items used in existing literature (Hassenzahl et al., 2003; Page & Herr, 2002). This is a natural result of sourcing descriptions of aesthetic pleasure in the literature to use as input into the research.

Additionally, since the factor loadings for the items measuring aesthetic pleasure were all very high (>.90), suggesting that each item measures the same construct approximately equally (Streiner, 2003), one might wonder whether it is necessary to use all five scale items in a future study. Literature is divided on whether multi-item or single-item scales are preferred in research. Studies have shown that the predictive validity of multi-item versus single-item scales varies between constructs. Multi-item scales show better predictive validity for more ambiguous constructs and/or stimuli, because the items each capture a separate facet of the construct they are intended to measure (Baumgartner & Homburg, 1996; Bergkvist & Rossiter, 2007). Single-item scales are often suitable for concrete and singular constructs (Rossiter, 2002) and are preferred for practical reasons (e.g., time constraints in a questionnaire; Bergkvist & Rossiter, 2007).

Conceptually, we can argue that aesthetic pleasure is concrete and singular; in the mind of the rater it is "easily and uniformly imagined" (Bergkvist & Rossiter, 2007, p. 176). On a practical level that would mean that a researcher could suffice with using only one item or, if preferred, only a few, and thus does not need to use all five items to measure the construct aesthetic pleasure. However, other researchers say that this is only appropriate if the construct is an observable construct and not a latent construct (e.g., buying behavior is observable, while attitudes are not; Jöreskog & Sörbom, 1979; MacCallum & Austin, 2000). If seen as such a latent construct, then multi-item scales are preferred. We argue that aesthetic pleasure, although uniformly defined, is a latent construct as it cannot be directly and objectively observed. Therefore, we argue that choosing more than one item would be best to capture the full construct of aesthetic pleasure. Moreover, choosing which item to use if a researcher wishes to use only one item can be problematic. Several ways of approaching this choice have been researched: face-validity value by researchers themselves (Bergkvist & Rossiter, 2007), by an expert panel (Rossiter, 2002) or on a statistical basis (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). All these have their problems: the first two are subjective, and the latter is objective, but choosing the item with the highest loading may be incorrect due to sampling bias (e.g., in another sample, another item could have the highest loading; Darden, Carlson, & Hampton, 1984). Therefore, we advise using several items of our scale (e.g., the three with the highest loading, or the three that make the most sense conceptually for the chosen stimuli) to be sure that the whole construct is captured for the sample and situation at hand.

In this research, we set out to develop a scale to measure aesthetic appreciation in design and therefore product designs and websites were used as stimuli. The use of designed artifacts in our research allowed us to capture aesthetic pleasure in its "pure" sense, because as outlined in the introduction, aesthetic pleasure is often the *only* aesthetic response people have to product designs, next to experiences

related to, for instance, affordances, usability, and expressive meaning (Hekkert & Schifferstein, 2008; Norman, 1988). Our scale, however, measures the immediate pleasure we attain from perceiving something "for its own sake" and is therefore not necessarily restricted to use in the context of design. Since it measures the aesthetic response as such, it could therefore also be used to capture the aesthetic pleasure of all kinds of other instances, whether they are a natural scene, a butterfly, a human face, a piece of architecture, or a painting by Van Gogh. We very much encourage studies in other diverse fields to further validate our scale and test its generalizability to domains other than designed artifacts.

Similarly, even though the items were validated using visual stimuli, we argue that these items can also be used to measure aesthetic pleasure following perception with other sensory modalities, and can even be applied to capture aesthetic responses resulting from more conceptual processing of objects. Accordingly, the items measuring aesthetic pleasure have already successfully been applied to assess the relationship of unity and variety with aesthetic pleasure in the tactile domain (Post, Blijlevens, & Hekkert, 2013), as well as for measuring the aesthetic pleasure people attain from understanding designer's intentions for the product design (Da Silva, Crilly, & Hekkert, 2015). Future research should also attempt to assess the generalizability of the scale to other instances that can be aesthetically appraised with the various senses, as well as to other conceptual phenomena.

Group comparisons showed that all the items used to measure aesthetic pleasure can be used in different languages. We did, however, notice differences in the individual items that were best at measuring the intended construct between countries. This can be due to translation issues; however, stringent translate/back translate methods were used to construct the items. Another explanation can be that certain words are more common in one language than another. On a practical level, this means that different combinations of items can be chosen to measure aesthetic pleasure depending on the country in which the research is to be performed. However, we expect that the choice should only make a marginal difference, because all factor loadings were very high for each item in each country.

It is intended that the development of this scale will enable meaningful comparisons between studies of aesthetics that will help to elucidate the relationships between aesthetic pleasure and its determinants. Furthermore, practitioners can use this scale to reliably assess the aesthetic pleasure induced by a creation, and can therefore be properly informed about the impact of their designs and the kind of factors underlying this response. As such, the research can ultimately have many practical implications for guiding designers and architects in creating aesthetically pleasing artifacts.

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research, 34,* 347–356. http://dx.doi.org/10.2307/3151897

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423. http://dx.doi.org/10.1037/0033-2909.103.3.411

Arbuckle, J. L. (1995). *Amos for Windows. Analysis of moment structures* (Version 3.5). Chicago, IL: SmallWaters.

Augustin, M. D., Carbon, C.-C., & Wagemans, J. (2011). Measuring aesthetic impressions of visual art. *Perception, 40,* 219.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13,* 139–161. http://dx.doi.org/10.1016/0167-8116(95)00038-0

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44,* 175–184. http://dx.doi.org/10.1509/jmkr.44.2.175

Blijlevens, J., Carbon, C.-C., Mugge, R., & Schoormans, J. P. L. (2012). Aesthetic appraisal of product designs: Independent effects of typicality and arousal. *British Journal of Psychology, 103,* 44–57. http://dx.doi.org/10.1111/j.2044-8295.2011.02038.x

Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *Journal of Marketing, 59,* 16–29. http://dx.doi.org/10.2307/1252116

Bullough, E. (1912). 'Psychical distance' as a factor in art and an aesthetic principle. *British Journal of Psychology, 5,* 87–118. http://dx.doi.org/10.1111/j.2044-8295.1912.tb00057.x

Creusen, M. E. H., & Schoormans, J. P. L. (2005). The different roles of product appearance in consumer choice. *Journal of Product Innovation Management, 22,* 63–81. http://dx.doi.org/10.1111/j.0737-6782.2005.00103.x

Cupchik, G. C., & Laszlo, J. (Eds.). (1992). *Emerging visions of the aesthetic process: Psychology, semiology, and philosophy*. New York, NY: Cambridge University Press.

Darden, W. S., Carlson, M., & Hampton, R. D. (1984). Issues in fitting theoretical and measurement models in marketing. *Journal of Business Research, 12,* 273–296. http://dx.doi.org/10.1016/0148-2963(84)90012-2

Da Silva, O., Crilly, N., & Hekkert, P. (2015). How people's appreciation of products is affected by their knowledge of the designers' intentions. *International Journal of Design, 9,* 21–33.

Desmet, P. (2003). Measuring emotion: Development and application of an instrument to measure emotional responses to products. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology* (pp. 111–123). Norwell, MA: Kluwer Academic. http://dx.doi.org/10.1007/1-4020-2967-5_12

Desmet, P. M. A., & Hekkert, P. (2007). Framework of product experience. *International Journal of Design, 1,* 13–23.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science, 40,* 434–449. http://dx.doi.org/10.1007/s11747-011-0300-3

Dutton, D. (2009). *The art instinct*. New York, NY: Oxford University Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50. http://dx.doi.org/10.2307/3151312

Frijda, N. H. (1989). Aesthetic emotions and reality. *American Psychologist, 44,* 1546–1547. http://dx.doi.org/10.1037/0003-066X.44.12.1546

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualitat [AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality]. In J. Ziegler & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: B. G. Teubner. http://dx.doi.org/10.1007/978-3-322-80058-9_19

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction, 25,* 235–260. http://dx.doi.org/10.1080/07370024.2010.500139

Hekkert, P. (2006). Design aesthetics: Principles of pleasure in design. *Psychology Science, 48,* 157–172.

Hekkert, P. (2014a). Aesthetic responses to design: A battle of impulses. In T. Smith & P. Tinio (Eds.), *The Cambridge handbook of the psychology of aesthetics and the arts* (pp. 277–299). New York, NY: Cambridge University Press.

Hekkert, P. (2014b). What I talk about when I talk about aesthetics. In A. Kozbelt (Ed.), *Proceedings of the 23rd Biennial Congress of the International Association of Empirical Aesthetics* (pp. 244–246).

Hekkert, P., & Schifferstein, H. N. J. (2008). Introducing product experience. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product experience* (pp. 1–8). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/B978-008045089-6.50003-4

Hekkert, P., Snelders, D., & Van Wieringen, P. C. W. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology, 94,* 111–124. http://dx.doi.org/10.1348/000712603762842147

Hoyer, W. D., & Stokburger-Sauer, N. (2012). The role of aesthetic taste in consumer behavior. *Journal of the Academy of Marketing Science, 40,* 167–180. http://dx.doi.org/10.1007/s11747-011-0269-y

Hung, W., & Chen, L. (2012). Effects of novelty and its dimensions on aesthetic preference in product design. *International Journal of Design, 6,* 81–90.

Jais, S. D. (2006). *The successful use of information in multinational companies—An exploratory study of individual outcomes and the influence of national culture*. Wiesbaden, Germany: Deutscher Universitäts Verlag.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426. http://dx.doi.org/10.1007/BF02291366

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Kant, I. (1952). *The critique of judgment* (J. C. Meredith, Trans.). Oxford, UK: Clarendon Press. (Original work published 1790)

Kreitler, H., & Kreitler, S. (1972). *Psychology of the arts*. Durham, NC: Duke University Press.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies, 60,* 269–298. http://dx.doi.org/10.1016/j.ijhcs.2003.09.002

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95,* 489–508. http://dx.doi.org/10.1348/0007126042369811

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—Developments and challenges in empirical aesthetics. *British Journal of Psychology, 105,* 443–464. http://dx.doi.org/10.1111/bjop.12084

Leder, H., Ring, A., & Dressler, S. G. (2013). See me, feel me! Aesthetic evaluations of art portraits. *Psychology of Aesthetics, Creativity, and the Arts, 7,* 358–369. http://dx.doi.org/10.1037/a0033311

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226. http://dx.doi.org/10.1146/annurev.psych.51.1.201

Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *The American Journal of Psychology, 103,* 53–80. http://dx.doi.org/10.2307/1423259

Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies, 68,* 689–709. http://dx.doi.org/10.1016/j.ijhcs.2010.05.006

Norman, D. (1988). *The design of everyday things*. New York, NY: Doubleday.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Page, C., & Herr, P. M. (2002). An investigation of the processes by which product design and brand strength interact to determine initial affective and quality judgments. *Journal of Consumer Psychology, 12,* 133–147. http://dx.doi.org/10.1207/S15327663JCP1202_06

Post, R. A. G., Blijlevens, J., & Hekkert, P. (2013). The influence of unity-in-variety on aesthetic appreciation of car interiors. In K. Sugiyama (Ed.), *Consilience and Innovation in Design: Proceedings of the 5th International Congress of International Association of Societies of Design Research* (pp. 1–6). Tokyo, Japan: Shibaura Institute of Technology.

Postrel, V. (2003). *The substance of style: How the rise of aesthetic value is remaking commerce, culture, and consciousness*. New York, NY: HarperCollins.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing, 19,* 305–335. http://dx.doi.org/10.1016/S0167-8116(02)00097-6

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819. http://dx.doi.org/10.1037/0022-3514.76.5.805

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal process in emotion: Theory, methods, research*. New York, NY: Oxford University Press.

Schoormans, J. P. L., & Robben, H. S. J. (1997). The effect of new package design on product attention, categorization and evaluation. *Journal of Economic Psychology, 18,* 271–287. http://dx.doi.org/10.1016/S0167-4870(97)00008-1

Spangenberg, E. R., Voss, K. E., & Crowley, A. E. (1997). Measuring the hedonic and utilitarian dimensions of attitude: A generally applicable scale. In M. Brucks & D. J. MacInnis (Eds.), *NA—Advances in consumer research* (Vol. 24, pp. 235–241). Provo, UT: Association for Consumer Research.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80,* 99–103. http://dx.doi.org/10.1207/S15327752JPA8001_18

Swami, V. (2013). Context matters: Investigating the impact of contextual information on aesthetic appreciation of paintings by Max Ernst and Pablo Picasso. *Psychology of Aesthetics, Creativity, and the Arts, 7,* 285–295. http://dx.doi.org/10.1037/a0030965

Tractinsky, N., Shoval-Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13,* 127–145. http://dx.doi.org/10.1016/S0953-5438(00)00031-X

Veryzer, R. W., Jr., & Hutchinson, J. W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of Consumer Research, 24,* 374–385. http://dx.doi.org/10.1086/209516

Whitfield, T. W. A., & Slatter, P. E. (1979). The effects of categorisation and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology, 70,* 65–75. http://dx.doi.org/10.1111/j.2044-8295.1979.tb02144.x

(*Appendices follow*)

Appendix A

Items Generated in Phase 1 of Item Generation Procedure

The 86 Items That Resulted From Extensive Literature Search in Phase 1 of the Item Generation Procedure

Aesthetic pleasure		Determinants	
Motivating	Joyful	Patchy	Grasping
Warm feeling	Thrills or chills	Perfection	Professionally made
Emotive	Pleasant	Conveys presence	Complexity
Sublime	Pleasurable	Inventive	Well Finished
Arousing	Appealing	Density	Appropriate
Memorable	Gratifying	Clear	Dynamic
Confers quality	Attractive	Averageness	Harmonic
Intense	Beautiful	Legibility	Understandable
Inspiring	Positive	Up-to-date	Categorizable
Care	Delightful	Designed	Meaningful
Relaxed		Special effects	Comprehensible
Exciting		Clean	Coherent
Touched		Convenient	Fluent to process
Moved		Easy orientation	Typical
Fascinating		Creative	Orderly
Inviting		Symmetrical	Easy to use
Aesthetic		Distinctive	Structured
Favorable		Elicits associations	Varied
Pretty		Familiar	Conceptual
Good		Novel	Elated
Preference		Goes together	Powerful
Interesting		Graspable	
Appreciating		Botched	
Nice		Sophisticated	
Like		Original	
Awe		Grabs attention	
Elation		Easy to navigate	

Appendix B

Items ranked in Phase 2 of Item Generation Procedure

The 37 Items That Were Ranked According to Their Appropriateness for Measuring Aesthetic Pleasure in Phase 2 of the Item Generation Procedure

1	2	3	4	5
Motivating	Relaxed	Inviting	Like	Pleasant
Warm feeling	Exciting	Aesthetic	Awe	Pleasurable
Emotive	Touched	Favorable	Elation	Appealing
Sublime	Moved	Pretty	Gratifying	
Arousing	Fascinating	Good	Attractive	
Memorable		Preference	Beautiful	
Confers quality		Interesting	Positive	
Intense		Appreciating	Delightful	
Inspiring		Nice	Joyful	
Care			Thrills or chills	

Note. 1 = *least appropriate to measure aesthetic pleasure*, 5 = *most appropriate to measure aesthetic pleasure*.

(*Appendices continue*)

Appendix C

Items used in Phase 3 of Item Generation Procedure and Mean Representative Ratings

The 23 Items Used in the Third Phase of the Item Generation Procedure and the Mean Ratings of How Representative These Items Are to Measure Aesthetic Pleasure

Items	Mean representative
Appealing	3.60
Thrilling	1.80
Aesthetic	3.20
Satisfying	2.40
Beautiful	4.40
Pretty	2.20
Attractive	3.40
Positive	2.00
Delightful	3.40
Pleasurable	3.60
Favorable	2.80
Good	2.60
Pleasant	3.00
Gratifying	3.00
Inviting	2.40
Nice	2.80
Joyful	1.80
Interesting	1.80
Prefer	3.00
Like	3.40
Elates me	2.60
Appreciate	3.00
Leaves me in awe	1.60

Appendix D

Final Items in English, Dutch and Mandarin

Items for Each Construct (Aesthetic Pleasure, Typicality, Novelty, Unity, and Variety) in the Languages English, Dutch, and Mandarin, Respectively (Example of Camera as Stimulus of Interest)

English	Dutch	Mandarin
Visually, . . .	Visueel gezien, . . .	在外觀視覺上
	Aesthetic pleasure	
. . . this is a beautiful camera	. . . is dit een mooie camera	這是一台漂亮的相機
. . . this is an attractive camera	. . . is dit een aantrekkelijke camera	這是一台具有吸引力的相機
. . . this camera is pleasing to see	. . . is deze camera prettig om te zien	這台相機看起來讓人感到愉快
. . . this camera is nice to see	. . . is deze camera aangenaam om naar te kijken	這台相機看起來不錯
. . . I like to look at this camera	. . . vind ik het fijn om naar deze camera te kijken	我喜歡注視這台相機
	Typicality	
. . . this is a typical camera	. . . is dit een doorsnee camera	這是一台典型的相機
. . . this is representative of a camera	. . . is dit representatief voor een camera	這是一台具有代表性的相機
. . . this design is common for a camera	. . . is dit ontwerp gangbaar voor een camera	這個設計對相機而言是常見的
. . . this is a standard design	. . . is dit een standaard ontwerp	這是一個標準的設計
. . . this is characteristic of a camera	. . . is dit kenmerkend voor een camera	這具有相機特徵
	Novelty	
. . . this is a novel camera	. . . is dit een nieuwe camera	這是一台新奇的相機
. . . this design is original	. . . is dit ontwerp origineel	這個設計是原創的

(*Appendices continue*)

Appendix D (*continued*)

English	Dutch	Mandarin
. . . this is a new example of a camera	. . . is dit een nieuw voorbeeld van een camera	這是一個新的相機案例
. . . this design is innovative	. . . is dit ontwerp innovatief	這個設計是創新的
	Unity	
. . . this is a unified design	. . . is dit een samenhangend ontwerp	這個設計是一致的
. . . this is an orderly design	. . . is dit een ordelijk ontwerp	這是一個整齊有序的設計
. . . this is a coherent design	. . . is dit een coherent ontwerp	這是個連貫有條理的設計
	Variety	
. . . this design is made of different parts	. . . bevat dit ontwerp verschillende onderdelen	這個設計是由不同元件形成的
. . . this design conveys variety	. . . drukt dit ontwerp variatie uit	這個設計傳達出多樣性
. . . this design is rich in elements	. . . is dit ontwerp rijk aan elementen	這個設計有豐富多元的元素

Received January 15, 2015
Revision received September 28, 2016
Accepted September 30, 2016 ■