KU LEUVEN

FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

ORDER, COMPLEXITY, AND AESTHETIC PREFERENCES FOR NEATLY ORGANIZED COMPOSITIONS

Master’s thesis submitted for the
degree of Master of Science in
Master of Psychology: Theory and
Research by
Eline Van Geert

Supervisor: Prof. Dr. Johan
Wagemans

2016-2017

KU LEUVEN

FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

ORDER, COMPLEXITY, AND AESTHETIC PREFERENCES FOR NEATLY ORGANIZED COMPOSITIONS

Master's thesis submitted for the degree of Master of Science in Master of Psychology: Theory and Research by
Eline Van Geert

Supervisor: Prof. Dr. Johan Wagemans

2016-2017

Summary

With more than 500.000 people following the Tumblr-blog of Things Organized Neatly© (http://thingsorganizedneatly.tumblr.com/), images of neatly organized compositions seem to be quite popular online. This raises the question why this type of images is so popular and attracts so many people.

To address this issue, we investigated which aspects of the images and their observers influenced aesthetic preferences for images of neatly organized compositions. We hypothesized the theoretical principle of balancing order and complexity to play an important role in determining aesthetic preferences, especially for this type of stimuli. Therefore, the selection of specific stimulus and person properties to study focused on factors relating to (the balance between) order and complexity. In a large-scale online study, 415 participants chose for each of 100 image pairs which one of two simultaneously presented images they preferred and completed some personality questionnaires (e.g., Personal Need for Structure). In a second (optional) part of the study, 84 participants also rated how ordered, complex, soothing, and fascinating they perceived each of 184 individual images to be. Additionally, some objective statistical measures were calculated on the images (i.e., PHOG-derived measures of self-similarity, complexity, and anisotropy, Fourier slope, and fractal dimension). Subsequently, we conducted correlation and regression analyses to investigate how and which stimulus and person properties related to (a) aesthetic preferences for neatly organized compositions; (b) judgments of order and complexity; and (c) judgments of soothingness and fascination.

Aesthetic preferences for images of neatly organized compositions relate positively with both the perceived soothingness and perceived fascination of the images. Images high in order and high in complexity were perceived as more fascinating. Images high in order and low in complexity were perceived as more soothing. As subjective order and complexity seem to have rather independent effects on positive aesthetic appreciation, the balance between order and complexity seems to be a combination of the independent effects of order and complexity rather than an interaction. Some objective indicators of complexity (i.e., HOG-based complexity, fractal dimension, and to a certain extent Fourier slope) related strongly positively with subjective complexity. Individuals differ in their preferences for more ordered, complex, soothing, and fascinating images, and these differences can partly be explained by differences in personality (i.e., symmetry, ordering, and arranging tendencies, personal need for structure, openness to experience) and age. In general, stimulus and person interact in determining aesthetic appreciation, and further studies should focus on empirically investigating and theoretically explaining these interactions.

Acknowledgements

In the first place, I want to thank my supervisor, Prof. Dr. Johan Wagemans, who assisted me wisely in designing the study and improving this manuscript. I want to thank Mr. Rudy Dekeerschieter for implementing my study online, Prof. Dr. Christoph Redies for calculating the statistical image properties for the images used in my study, and Prof. Dr. Eva Ceulemans and Dr. Pieter Moors for their statistical advice. For data collection and for personal and emotional support, I am happy that I could count on my family and friends, for whom it might not always have been easy to deal with me during times of high stress.

Contribution and approach

For my master's thesis I independently reviewed and synthesized the literature on aesthetic appreciation and the balance between order and complexity. Although I got a paper to start from, I collected most papers for the literature review myself. Additionally, my supervisor provided me with many interesting literature tips during the two years I was working on my thesis, which I tried to integrate in my review. With the help and advice of my supervisor, I created my own study design, hypotheses, and analysis plan, collected and selected the necessary materials (i.e., images, questionnaires, etc.), and applied for ethical application of the study. For the online implementation of my study, I am happy that I could count on Mr. Rudy Dekeerschieter, who magically turned my study implementation plan into reality. Besides creating an advertisement for my study, spreading it around Flanders and abroad, and asking everyone I could think of to participate, the online data collection was heavily supported by my family and friends, who I am very grateful. Without them, it would not have been possible to collect such a large and diverse sample of participants. I independently processed all the gathered data and analyzed them using R. I want to thank my supervisor for re-, re- and reviewing this manuscript, reading the plenty of questions and comments I left in the margin, and giving me the necessary confidence to develop myself as an independent researcher.

Besides this manuscript, I also disseminated my project in other ways. By the time of my thesis defense, I will have presented my work twice on the annual meeting of the Belgian Association for Psychological Science. The first year I presented a poster on my theoretical work and study design, in this second year I will give an oral presentation about my study results. Additionally, I participated in the Open Science Day ("Dag van de Wetenschap") to present my project to a wider public (and also to gather some more participants), and presented my project twice in the lab meetings of the Laboratory for Experimental Psychology at KU Leuven.

Although my manuscript is quite long, I think the length is justified by the "complexity" (to stay in the topic!) of the research problems I am working on and of the conducted study (i.e., various dependent and independent variables to investigate, several possible units of analysis). I hope I could also provide you with the necessary "order" to compensate the complexity of the discussed material. But as the lead singer of OK Go, Damien Kulash, stated it so nicely in an interview with The Washington Post[1]: "[…] I would say that in general simplicity is really complicated".

[1] Ohlheiser, A. (2016, December 13). After 10 years of viral videos, what OK Go has learned about the Internet. *The Washington Post.* Retrieved from https://www.washingtonpost.com/

Table of Contents

List of Figures

Order, Complexity, and Aesthetic Preferences for Neatly Organized Compositions

With more than 500.000 people following the Tumblr-blog of Things Organized Neatly© (http://thingsorganizedneatly.tumblr.com/) curated by Austin Radcliffe, images of neatly organized compositions seem to be quite popular online (Radcliffe, 2016). This raises the question why this type of images is so popular and attracts so many people. To address this issue, we investigated what aspects of the images and their spectators underlie people's experience of neatly organized compositions as aesthetically pleasing. As we hypothesized the balance between order and complexity to play an important role in determining aesthetic preferences, especially for this type of stimuli, the selection of specific stimulus and individual difference dimensions to include in the study focused on factors relating to (the balance between) order and complexity. These factors were then tested in a large-scale online study with carefully selected stimuli.

Neatly Organized Compositions

A neatly organized composition is a set of objects, or parts of objects, organized in an orderly (i.e., tidy, neat) way (see Figure 1 for examples). Other terms that are related to this topic are organization porn (i.e., stylized images of everyday objects arranged in a neat, visually pleasing way; Alleyne, 2015), knolling (i.e., arranging objects in parallel or in right angles as a way to organize them, a term first used by Andrew Kromelow in 1987 and popularized by Tom Sachs in 2009; Hay, 2015; Sachs, 2011), and flat lay photography (i.e., objects spread out on a flat surface and photographed from above; Innis, 2016). Although the blog of Things Organized Neatly© is a popular reference (Alleyne, 2015), images of neatly organized compositions circulate on various online sharing platforms like Reddit, Pinterest and Tumblr as well as on social media like Instagram (Alleyne, 2015). Ursus Wehrli, a Swiss artist, is also working with neatly organized compositions, both of daily-life objects and pieces of art (cf. books like “Tidying Up Art” and “The Art of Clean Up”; Wehrli, 2002, 2011). Wehrli even created an app to let people tidy up art themselves (Kein & Aber AG, 2012). Furthermore, his TED talk about tidying up art given in 2006 but put online in 2008 has more than 1.250.000 views (Green, 2013; Wehrli, 2006).

In lay interpretations, the appreciation of neatly organized images is often linked to obsessive-compulsive tendencies (e.g., Bielski, 2010; The Webbys, 2015). Furthermore, lay interpretations emphasize that some people watch the neatly organized images to calm down or to relax (e.g., Alleyne, 2015; Ellison, 2015), as these images bring “a sense of order to the chaos of everyday life” (Ellison, 2015).

Understanding Aesthetic Appreciation

The popularity of neatly organized compositions, especially online, urges us to think about reasons why people experience them as aesthetically pleasing. This project can be situated in the field of experimental psycho-aesthetics, as it is aimed at accurately describing people's aesthetic evaluations and understanding the factors that determine them (Palmer, Schloss, & Sammartino, 2013; Wagemans, 2011). We will use aesthetic *appreciation* to refer to the taste or sense of beauty (or ugliness) based on feelings of pleasure (or displeasure) attained from the sensory processing of a stimulus “for its own sake” (Post, Blijlevens, & Hekkert, 2016; Wagemans, 2011). This definition is closely related to Baumgarten’s definition of aesthetics (Wagemans, 2011). Aesthetic *preference* will be used to indicate observers’ more positive aesthetic appreciation for a certain stimulus in contrast to the other stimuli presented.

Figure 1. Examples of neatly organized compositions.

Aesthetics is not confined to art. It is important to clarify that aesthetics is not reducible to art or the other way around (Palmer et al., 2013). Whereas aesthetic appreciation can occur in response to very diverse types of stimuli, not constrained to art objects, art pertains to more than people’s aesthetic appreciations in response to the artwork (Palmer et al., 2013). Palmer et al. (2013) argue that essentially everyone has some aesthetic response to essentially everything they see. In most situations this response may not be on the forefront of attention, but it can be under certain conditions, for example when the aesthetic response is extreme (Palmer et al., 2013).

Images of neatly organized compositions in particular are a type of stimuli situated on this border between art and aesthetics of every-day objects and images. Whereas some people will view them as a type of art, others might be more sceptical of these images as an art form. In conclusion, regardless of whether people view neatly organized images as art, they might experience a form of aesthetic appreciation in response to this type of stimuli.

Aesthetics involves both stimulus and person. Research in experimental psycho-aesthetics can focus on either stimulus or person properties determining aesthetic preferences (Wagemans, 2011), but an adequate theory of aesthetic preferences must include both, from an interactionist perspective (Mather, 2014). Although most studies of aesthetic preference aim to find stimulus properties influencing preferences of all individuals in the same way (Vessel & Rubin, 2010), the influence of individual differences in aesthetic preference must not be neglected. Previous research has shown that individual differences in aesthetic preferences can be strong and that ignoring these by averaging over participants can lead to an inadequate representation or an incomplete interpretation of the acquired data (Güçlütürk, Jacobs, & van Lier, 2016; Jacobsen, 2004; Vessel & Rubin, 2010). Vessel and Rubin (2010) found more pronounced individual differences in aesthetic preferences for abstract than for real-world images. However, their experiments also indicated that pronounced differences in preference can occur for real-world images in contexts in which the semantic meaning of the images is de-emphasized. In the case of images of neatly organized compositions, the organization brought into these images could de-emphasize the semantic meaning of the presented objects. Therefore, we expect that pronounced individual differences in preference exist for images of neatly organized compositions. An important and defining aspect of the current project is that both aspects of the stimuli and of the individuals seeing the stimuli are investigated.

Balancing ecological validity and experimental control. Research in experimental psycho-aesthetics often struggles with the challenge of balancing the needs for ecological validity and experimental control. In studies investigating the influence of order and complexity on aesthetic appreciation, maximizing experimental control by using simple stimuli has long been the dominant approach (Wagemans, 2011). This approach gives researchers the possibility to precisely control which factor is manipulated and it thereby enables them to determine which factor is responsible for the difference in aesthetic appreciation. However, the stimuli used are often that simple that they are not very relevant to aesthetic objects and images encountered in daily life, and could therefore neglect important determinants of real aesthetic behavior (Nadal, 2007). In contrast, other studies try to maximize ecological validity by using existing aesthetic stimuli, like art works or design objects (Wagemans, 2011). Although this increases ecological validity, it also diminishes experimental control and thereby the possibility to precisely determine the crucial differences between the different stimuli presented (Nadal, 2007). In this first study concerning images of neatly organized compositions, we tried to balance the needs for ecological validity and experimental control by collecting existing images but only pairing and comparing images that seemed to differ on few stimulus dimensions and sometimes slightly adapting the existing images (cf. Methods section and Appendix A).

Behavioral measures of aesthetic appreciation. Gustav Theodor Fechner (1834-1887), pioneer of experimental aesthetics (e.g., Gilbert & Kuhn, 1953/1972), described three distinct experimental methods for investigating aesthetic appreciation: the method of construction or production, the method of selection or choice, and the method of use (Fechner, 1876; Gilbert & Kuhn, 1953/1972; Nadal, 2007; Westphal-Fitch, Oh, & Fitch, 2013). When using the *method of production*, participants are asked to produce their own aesthetic objects. The *method of use* is concerned with measuring objects in daily life and analyzing patterns in these observational data, under the assumption that the features that are most common in a certain society are also preferred by that society. The *method of selection or choice* was the one Fechner used the most in his experiments and that is still frequently used in contemporary research. It involves asking observers to compare the pleasingness of different stimuli. A common used form of the method of choice is two-alternative forced choice (2AFC). In spatial (or temporal) 2AFC, participants indicate which of two simultaneously (or sequentially) presented visual stimuli they prefer. Advantages of this method include that there is virtually no memory load, that the response asked from the participant is very simple, and that minimal response bias effects occur (Palmer et al., 2013). However, 2AFC data give only relative rather than absolute information about the aesthetic appreciation of certain types of stimuli. In other words, choice data can tell which stimulus people prefer over others, but it cannot tell the extent to which people judge a stimulus as ugly or beautiful (Mather, 2014). An alternative method that remedies this disadvantage is showing participants one stimulus at a time and asking them to rate this stimulus on a scale of aesthetic appreciation (e.g., rate how much they like the stimulus on a 7-point scale). However, this rating method has its own disadvantages. For example, it can be hard for participants to make consistent ratings across trials, especially at the beginning of the series. Therefore, when using this rating method, it is suggested to present the entire set or a representative sample of stimuli before the participants start rating (Palmer et al., 2013). In this study, we chose for a spatial 2AFC method to measure aesthetic preferences for neatly organized compositions.

Complexity and Aesthetic Appreciation

One factor assumed to play an important role in determining aesthetic appreciation is the complexity of a stimulus. Arnheim (1966) defined complexity as “the multiplicity of the relationships among the parts of an entity” (p. 123). As will become clear during my literature review on the topic, the term complexity has been used to refer to a whole plethora of different stimulus aspects and has been defined in very diverse ways. Therefore, I will discuss the definition of this concept more elaborately in a later section.

Differential relationships between complexity and aesthetic appreciation. Although a lot of studies support the importance of complexity in shaping aesthetic preferences (e.g., Jacobsen & Höfel, 2002; Jacobsen, Schubotz, Höfel, & von Cramon, 2006; Tinio & Leder, 2009), studies do not agree on the type and direction of the relationship (Nadal, Munar, Marty, & Cela-Conde, 2010). Berlyne predicted that stimuli with a moderate level of complexity would be preferred above those with low or high levels of complexity (Nadal, 2007). Many studies support Berlyne’s hypothesis of an inverted U-shaped relationship between complexity and aesthetic preference (e.g., for music: North & Hargreaves, 1995; Gordon & Gridley, 2013; for line drawings of house facades: Imamoglu, 2000; for language sequences and random shapes: Munsinger & Kessen, 1964). However, other studies reported a positive linear (e.g., for snowflakes: Adkins & Norman, 2016; for texture patterns: Friedenberg & Liby, 2016), or even a non-inverted U-shaped relationship (e.g., for solid objects: Adkins & Norman, 2016; Philips, Norman, & Beers, 2010) between complexity and aesthetic appreciation.

The contradictory findings seem to result from several theoretical and empirical difficulties (Nadal et al., 2010). Not only do studies differ enormously in the type of stimuli they used (as well as in the number of participants, the number of stimuli and the ecological validity of these stimuli; Marin & Leder, 2013), also the way in which complexity was defined, measured, and manipulated varied between studies (Nadal et al., 2010). Additionally, the effects of complexity on aesthetic appreciation might be subject to individual differences (e.g., Aitken, 1974; Güçlütürk et al., 2016; Jacobsen & Höfel, 2002; Tinio & Leder, 2009).

Objective and subjective complexity. One important distinction between different types and operationalizations of complexity is the distinction between subjective and objective complexity. *Objective complexity* refers to the amount or degree of complexity that is physically present in a certain stimulus. Consequently, objective measures of complexity are computed based on the properties of the stimulus itself. Whereas some objective measures are based on statistical image properties or image compression techniques, others are based on more directly visible stimulus dimensions. Commonly used examples of statistical image properties related to complexity are the Pyramid of Histograms of Orientation Gradients (PHOG) based measures of self-similarity, complexity, and anisotropy, Fourier slope and fractal dimension (e.g., Braun, Amirshahi, Denzler, & Redies, 2013; Mather, 2014). Popular measures based on image compression techniques include GIF (Graphics Interchange Format; e.g., Forsythe, Nadal, Sheehy, Cela-Conde, & Sawey, 2011; Friedenberg & Liby, 2016; Marin & Leder, 2013) and JPEG (Joint Photographic Expert Group; e.g., Chikhman, Bondarko, Danilova, Goluzina, & Shelepin, 2012; Marin & Leder, 2013). Objective complexity measures based on more directly visible characteristics of the stimulus may include different measures of the number and variety of elements or colors in the stimulus (e.g., number of individual elements; Tinio & Leder, 2009; number of independent turns in polygonal shapes; Munsinger & Kessen, 1964).

In contrast, *subjective complexity* entails the subjects' perception of the complexity of the stimulus in question. Consequently, measures of subjective complexity take the observer into account, but vary in how much importance they give to individual differences (i.e., averaging over participants or using the separate scores of each individual participant). Examples of how subjective complexity can be measured include 2AFC tasks (e.g., Chipman & Mendelson, 1979) and rating scales (e.g., Marin & Leder, 2013), in parallel to the measures of aesthetic appreciation discussed above (cf. "Behavioral measures of aesthetic appreciation").

Statistical image properties as objective measures of complexity. The objective complexity measures discussed here are the ones that are included in the present study. *Self-similarity* indicates how similar the object as a whole is to its parts (Lyssenko, Redies, & Hayn-Leichsenring, 2016). It is closely related to scale invariance and fractal dimension, and earlier studies have found museum paintings to be highly self-similar (e.g., Redies, Amirshahi, Koch, & Denzler, 2012). *HOG-based complexity* is defined as the sum of the magnitude of changes in luminance or color in an image. The higher the value of this measure, the more objectively complex the image is (Redies et al., 2012). *Anisotropy* measures the difference in magnitude of changes in luminance or color across orientations in an image (Braun et al., 2013). Low anisotropy indicates that the strength of the changes is similar across orientations, whereas high anisotropy indicates that some orientations are more prominent than others in the image. In previous research, museum paintings and graphic artworks were characterized by a low degree of anisotropy (e.g., Redies et al., 2012). *Fourier slope* is an indicator of the strength of low spatial frequencies (representing coarse detail) relative to high spatial frequencies (representing fine detail) in the image (Redies, Brachmann, & Hayn-Leichsenring, 2014). A slope value of -2 indicates that the image has fractal-like, scale-invariant properties, which means that the relative strength of low and high spatial frequencies stays constant when zooming in or out of the image. In images with a shallower slope (values higher than -2), high spatial frequencies are more prominent than in image with a slope of -2. In images with a steeper slope (values lower than -2), low spatial frequencies are more important (Redies et al., 2014). For images of Western graphic art and natural scenes, the Fourier slope was found to be around -2, indicating scale-invariant properties (e.g., Redies, Hasenstein, & Denzler, 2007). *Fractal dimension* indicates how ‘fractal’ or self-similar an image is (Hagerhall, Purcell, & Taylor, 2004). Fractals are shapes that show the same structures or patterns when zooming in or out of the image (Hagerhall et al., 2004). For two-dimensional images, the fractal dimension values are between 1 and 2, with values closer to 2 indicating more complex and intricate images (Mureika & Taylor, 2013). In previous research, humans preferred intermediate values for the fractal dimension, ranging from 1.3–1.5 (Taylor, Spehar, Van Donkelaar, & Hagerhall, 2011).

Although the relations between the different statistical image properties depend on the type of images analyzed (C. Redies, personal communication, July 2, 2016), earlier research consistently found a relatively high correlation between HOG complexity and fractal dimension (e.g., $r = .82$, $p < .001$ in Braun et al., 2013). The other correlations seemed more variable (for an idea of previously reported correlations, see Table 6 in Braun et al., 2013). Fourier slope and HOG complexity are somewhat related because a more shallow Fourier slope indicates more high frequencies (representing fine detail) in the image, which can make the image more complex. Anisotropy does only relate to complexity in some cases, for example in the most complex and self-similar natural images (which have low anisotropy; C. Redies, personal communication, July 2, 2016).

Aspects of objective complexity determining subjective complexity. Several researchers have tried to determine the physical and statistical (i.e., objective) aspects of the stimulus that help to predict subjective complexity (e.g., Attneave, 1957; Corchs, Ciocca, Bricolo, & Gasparini, 2016). In previous studies, statistical image properties like self-similarity and HOG complexity showed strong positive correlations with subjective ratings of complexity (e.g., $r = .56$ and $r = .68$, *p*s $< .001$, respectively, in Lyssenko et al., 2016). Anisotropy showed a negative correlation with subjective complexity ($r = –.39$, $p < .001$ in Lyssenko et al., 2016)[1].

However, subjective complexity is not reducible to aspects of objective complexity (Arnheim, 1954/2004; Attneave, 1957). Individual differences exist in which aspects of objective complexity determine an individual’s perceived complexity and to what extent (cf. “Individual differences in the perception of order and complexity”). Furthermore, the importance of different aspects of objective complexity for subjective complexity may vary depending on the type of stimuli selected (e.g., Chikhman et al., 2012) or on the definition of complexity given to the participants (e.g., Oliva, Mack, Shrestha, & Peeper, 2004).

Other researchers have specified a number of dimensions to categorize the different objective complexity factors influencing subjective complexity (e.g., Berlyne et al., 1968; Nadal et al., 2010). According to Berlyne (1960), the subjective or perceived complexity of a stimulus is positively related to the number of distinguishable elements and the dissimilarity between those elements. Additionally, he postulated a negative relation between perceived complexity and the degree to which several elements are responded to as a unit (Berlyne, 1960). A factor analysis conducted by Berlyne, Ogilvie, and Parham (1968) indicated two main factors determining subjective complexity: (a) the number of independently selected component elements, which they called an ‘information content’ dimension; and (b) a dimension they called ‘unitariness vs. articulation into easily recognizable parts’.

Chipman (1977) distinguished between quantitative and structural variables influencing perceived complexity. She argued that where the quantitative variables set the upper limit for the amount of perceived complexity, the detection of psychologically relevant organization (i.e., the structural variables) can reduce subjective complexity.

Nadal et al. (2010) later differentiated between three different forms of visual complexity that influenced people’s perception of complexity: subjective complexity related to (a) the amount and variety of elements, (b) the ways those elements are organized, and (c) asymmetry[2]. They also reported preliminary evidence for differential relationships between these three factors determining subjective complexity and aesthetic appreciation. Whereas the amount and variety of elements had a positive linear relationship with aesthetic appreciation, a descending U-shaped relation was found for complexity in the ways the elements in a stimulus are organized. Furthermore, asymmetry showed an inverted U-shaped relationship with aesthetic appreciation.

---

[1] Additionally, self-similarity and HOG complexity correlated slightly negatively and anisotropy slightly positively with structure ratings ($r = –.24$, $r = –.29$, and $r = .22$, ps $< .001$, respectively, in Lyssenko et al., 2016).

[2] Although the work of Nadal et al. (2010) is an important theoretical contribution for distinguishing between and structuring of the different uses of the term complexity in the existing literature, we will reinterpret these different types of complexity below in terms of order and complexity.

Complexity as a multidimensional concept. Besides Berlyne et al. (1968), Chipman (1977), and Nadal et al. (2010), other authors have also noted that visual complexity is a multidimensional concept (e.g., Kreitler, Zigler, & Kreitler, 1974; Rump, 1968). Rump (1968) reported that individuals' preferences for different aspects of objective complexity did not correlate with each other. This means that individuals' preference for a specific type of objective complexity did not reflect whether they also preferred other types of objective complexity. Kreitler et al. (1974) did find some correlations between the preferences for different complexity dimensions they examined, but the correlations were very low. However, most research on complexity and aesthetic appreciation has used complexity as a unidimensional concept, not distinguishing between different aspects of complexity (Nadal, 2007). As mentioned earlier, various operationalizations of complexity have been used, which may be one of the reasons why differential relationships between complexity and aesthetic appreciation have been found (Nadal, 2007). As Nadal et al. (2010) found preliminary support for differential relationships between the types of complexity they distinguished and aesthetic appreciation, it appears that acknowledging the multidimensionality of complexity may be a fruitful approach in the search for consistent relationships between complexity and aesthetic appreciation.

Subjective complexity resulting from objective complexity and order. The literature reviewed above (e.g., Berlyne et al., 1968; Nadal et al., 2010) defined subjective complexity as dependent on aspects related to the quantity and variety of elements in a stimulus as well as aspects related to the structure and organization in the stimulus. We agree with the authors mentioned above that subjective complexity may be defined based on these different dimensions. However, we will define *objective complexity* as only containing those aspects related to the quantity and variety of information in a stimulus, very similar to the first dimension mentioned by Berlyne et al. (1968) and Nadal et al. (2010). *Objective order* will be used to refer to aspects related to the structure and organization of the information in the stimulus (cf. the second dimension of Berlyne et al., 1968, and the second and third dimension of Nadal et al., 2010).

Order and Aesthetic Appreciation

Besides complexity, also the order or organization present in a stimulus is assumed to influence aesthetic appreciation. Arnheim (1966) defined order as "the degree and kind of lawfulness governing the relations among the parts of an entity" (p. 123). Although Nadal et al. (2010) described how the elements in the stimulus are organized and the symmetry present in the stimulus as aspects of complexity, we will regard them as aspects of order instead.

Simplicity versus order. It is important to clarify that we distinguish order from simplicity (and complexity from disorder). In what follows, simplicity is defined as opposite to complexity, indicating the quantity and variety of information content in a stimulus. In contrast, order refers to how and to what extent the elements in the stimulus are organized or structured, without taking into account the number or variety of elements that is present in the stimulus. Of course, the degree of order or disorder that can be present on a certain stimulus dimension will depend on the degree of complexity present on that dimension, in the sense that there is more room for an effect of order versus disorder when the number or variety of elements is larger. A stimulus that is complex on a certain stimulus dimension (e.g., showing a lot of variety in the orientation of its elements or its colors; see Figure 1A and 1D) can also be very ordered, both on the same dimension (e.g., a systematic gradient in colors can be presented; see Figure 1D) and/or on other stimulus dimensions (e.g., the elements can together form an ordered configuration, for example a circle; see Figure 1A). On the other hand, a stimulus that is very simple on a certain stimulus dimension (e.g., all elements in the stimulus have the same color; see Figure 1A) usually cannot be very disordered on the same dimension, but the stimulus can be very disordered on other stimulus dimensions (e.g., different orientations and shapes of the elements in the stimulus; see Figure 1A).

As was the case with complexity, different conceptualizations of order have also been used. However, to our knowledge, less systematic research is done about the different conceptualizations and operationalizations of order than those of complexity. Furthermore, whereas the distinction between objective and subjective order would make sense, it is not typically been used in the literature. One aspect of order that did get some empirical attention in previous decades when studying aesthetic appreciation is symmetry. One of the reasons that other aspects of order have remained understudied may be that those aspects seem less straightforward to measure in a purely objective way than symmetry or aspects of complexity. Furthermore, some aspects of order have been studied under the label of complexity (which has added to the confusion in the literature).

Order as a multidimensional concept. Like complexity, order can be viewed as a multidimensional concept. Different types of order exist (e.g., similarity grouping, symmetry, alternation/iteration, systematic alteration/gradient, etc.) and order can occur with respect to different dimensions (e.g., color, shape, spatial composition/configuration, etc.). Below some of the important components of order and their relation to aesthetic appreciation will be highlighted.

Perceptual grouping principles. Gestalt psychologists formulated different grouping laws (Arnheim, 1954/2004; for a review, see Wagemans et al., 2012) including the principles of similarity (e.g., in color, size, shape, or spatial orientation), proximity (i.e., similar spatial location), common fate (i.e., similar direction or speed), symmetry and parallelism, continuity, and closure. The most general Gestalt law is the law of Prägnanz, which states that objects within the perceptual field (and the perceptual field in itself) will be organized in the simplest and most encompassing way possible given the current conditions (Wagemans et al., 2012). In his book “Art and Visual Perception”, Arnheim (1954/2004) described the close relationship between the Gestalt principles and the arts: “at no time could a work of art have been made or understood by a mind unable to conceive the integrated structure of a whole” (p. 5) and “All works of art have to be looked at ‘from above,’ that is, with a primary grasp of the total organization. At the same time, however, relations among the parts often play an important compositional role” (p. 88).

Some authors hypothesized that aesthetic appreciation is generated from being able to group elements together and to detect the properties that bring order and unity to them (Post et al., 2016; Ramachandran & Hirstein, 1999). Also Koffka already described a link between Gestalt psychology and aesthetic appreciation (Koffka, 1940). However, to our knowledge, not much empirical research has systematically investigated the grouping principles above in relation to aesthetic appreciation.

Symmetry and aesthetic appreciation. One grouping principle for which evidence does exist, is symmetry. In the study of Jacobsen and Höfel (2002), symmetry was found to be the most important factor in the aesthetic appreciation of graphic patterns. Most studies argue that people prefer symmetry above non-symmetry (e.g., Cárdenas & Harris, 2006; Jacobsen et al., 2006; Munsinger & Kessen, 1964; Westphal-Fitch, Huber, Gómez, & Fitch, 2012). Cárdenas and Harris (2006) showed that participants preferred symmetrical designs above asymmetrical designs. Westphal-Fitch et al. (2012) used the method of production and found that people spontaneously produced highly ordered, often symmetrical, visual patterns, without being instructed to do so. However, some other investigations point to a preference for slight asymmetry, for example when comparing Mondrian's paintings to symmetrical alternatives (Swami & Furnham, 2012).

Perceptual balance and aesthetic appreciation. Symmetry can be seen as the simplest type of perceptual balance (Wilson & Chatterjee, 2005). Besides for the importance of symmetry, there is evidence for the importance of dynamic balance in determining aesthetic appreciation (Wilson & Chatterjee, 2005). Dynamic balance arises from a stimulus in which the visual forces of the different elements compensate each other (Wilson & Chatterjee, 2005). Also Hübner and Fillinger (2016) found perceptual balance to be a predictor for ratings of aesthetic appreciation. Locher, Stappers, and Overbeeke (1998) asked participants to create visual designs themselves. For most participants, the center of the design was very near to the geometric center of the display field, indicating perceptual balance.

Conceptual Dimensions of Order and Complexity

Although this was not explicitly discussed until now, order and complexity can also occur on conceptual or semantic dimensions (i.e., dimensions relating to concepts or ideas or to meaning). Birkhoff (1932/1933) distinguished formal and connotative elements of order. Formal elements of order (e.g., repetition, similarity, contrast, equality, symmetry, balance, and sequence) involve a reference to a simple physical property of the aesthetic stimulus. Connotative elements of order involve all properties of the stimulus that are not of this simple formal type. Nicki and Moss (1975) distinguished between perceptual and conceptual complexity, the first being related to the number of stimulus elements and the degree of irregularity in their arrangement and the second factor being related to the cognitive labels or associations evoked by the stimulus.

The fact that most studies in aesthetics have focused either on form (i.e., perceptual dimensions) or on content (i.e., conceptual or semantic dimensions) was mentioned by Berlyne (1960) as one of the problems with the aesthetics research of his time. Phillips et al. (2010) suggested that the interaction of 'denotative' and 'connotative' properties of artworks could drive the contrasting findings in the aesthetics literature in the sense that the connotations given to an artwork can be very peculiar for each observer, possibly leading to large individual differences.

Some studies found conceptual or semantic factors to be the most important ones in determining aesthetic appreciation (e.g., Martindale, Moore, & Borkum, 1990). In addition, some studies found evidence for a preference for less conceptual ambiguity. Stephens and Hoffman (2016) reported that textures with fewer descriptions (i.e., fewer possible conceptual organizations of the stimulus content) were liked better. Taylor and Franklin (2012) observed the same for colors: Colors associated with fewer objects were liked better than objects associated with many objects.

Whereas the results of Munsinger and Kessen (1964) supported the inverted U-shaped function between variability (i.e., complexity) and aesthetic appreciation, they found unexpectedly high preference for extremely simple and extremely complex figures. The preferences for extremely complex figures could be accounted for by effects of meaningfulness: The extremely complex figures were judged much more meaningful than the figures with intermediate complexity (Study 3). Also conceptual or semantic aspects of order and complexity seem to influence aesthetic appreciation.

Aesthetic Appreciation and the Balance Between Order and Complexity

Both order and complexity are thus important in determining aesthetic appreciation. One additional reason for the contrasting findings in both fields of study could be that effects order and complexity are intertwined, and that it is the balance between the two that influences aesthetic appreciation. We hypothesized the theoretical principle of balancing order and complexity to play an important role in explaining aesthetic preferences, especially for neatly organized compositions.

History of the balance between order and complexity. Throughout history, the importance of an appropriate balance between order (also referred to as unity, uniformity, synthesis, harmony, lawfulness, and organization) and complexity (also variety, diversity, or multiplicity are used) in explaining aesthetic appreciation has been suggested by many different philosophers, including Philo, Plotinus, Descartes, Hutcheson, Baumgarten and Moses Mendelssohn (Gilbert & Kuhn, 1953/1972; Rist, 1967). By introducing an aesthetics "from below"[3], Fechner's work represented a transformation from more deductive and theoretical to more inductive and experimental methods in the field of aesthetics (Cupchik, 1986). With his formal principle of unified connection of the manifold (Gilbert & Kuhn, 1953/1972), Fechner proposed that stimuli that are experienced as pleasing must provide an adequate balance between order (i.e., 'unified connection') and complexity (i.e., 'the manifold'; Cupchik, 1986). Furthermore, he argued that people will tolerate a moderate level of arousal more frequently and for a longer time than a very low or very high level, as this would cause under- or overstimulation, respectively (his 'principle of the aesthetic middle'; Cupchik, 1986).

[3] This is used in contrast to the philosophical aesthetics "from above" and not used in contrast to the earlier mentioned "from above" in the section about perceptual grouping principles.

Based on Wundt's (1874) inverted-U curve showing that pleasantness is perceived as highest for medium stimulus intensities, Berlyne proposed a similar association between pleasure and arousal potential: "moderate arousal potential will be maximally rewarding" (Berlyne, 1960, p. 201). However, this optimal level of arousal is supposed to depend on a large number of different factors, including collative (related to the viewer's expectations), psychophysical (related to the sensory dimensions of the stimulus), and ecological variables (related to the meaningfulness and associations to environmental objects; Palmer et al., 2013).

When applying his arousal theory to aesthetic appreciation, also Berlyne (1960) discusses the principle of unity in diversity. He relates diversity to two factors that constitute complexity (i.e., heterogeneity and numerosity of elements), but argues that diversity can also entail novelty, ambiguity, and surprise. In contrast, the principle of unity is proposed to moderate or diminish arousal levels (Berlyne, 1960).

Clarifying the relation between order and complexity. As long as no clear definition of unity (i.e., order), diversity (i.e., complexity), or their relation (i.e., balance) is given, we cannot really do anything with this information (Eysenck, 1942). Birkhoff (1932/1933), a mathematician, was the first who tried to define these concepts and their relation in a more exact way (Eysenck, 1942). His theory of aesthetic preference (M = O/C) stated that a measure of aesthetic preference (M) should vary positively with the amount of order (O) and negatively with complexity (C; Boselie & Leeuwenberg, 1985; Palmer et al., 2013). Birkhoff specified in detail how he defined order and complexity (i.e., identifying different stimulus properties that could be seen as different manifestations of order and complexity) for a series of different object classes, including polygons and melodies (Boselie & Leeuwenberg, 1985). He argued that the elements determining order and complexity should be identified for each class of objects separately (Eysenck, 1942). However, no convincing empirical evidence was found to support Birkhoff's formula (Boselie & Leeuwenberg, 1985; Palmer et al., 2013).

Eysenck (1942) acknowledged Birkhoff's contributions, but pointed out that next steps to take were (a) to identify the elements that determine order and complexity empirically (instead of via deductive reasoning); and (b) to derive a general formula to represent the relation between order and complexity in an experimental way. Therefore, Eysenck (1941) developed an empirical aesthetic formula to predict the preference of humans for simple polygons (Boselie & Leeuwenberg, 1985; Nadal, 2007). In Eysenck's measure (M = O x C) both order and complexity could positively contribute to aesthetic preference, whereas Birkhoff's measure predicted complexity to be negatively correlated with liking. Although Eysenck's measure specifically concerned the aesthetic preference for polygons, he predicted that the presence of aspects of order and complexity would be positively related to aesthetic preference in any set of visual stimuli (Nadal, 2007). However, he argued that the final formula would inevitably be more complicated than his proposal (M = O x C; Eysenck, 1942).

Arnheim postulated an antagonistic but also a complementary relationship between both concepts. More specifically, he stated that although order tends to reduce complexity and complexity tends to reduce order, order and complexity also need each other, as “complexity without order produces confusion; order without complexity causes boredom” (Arnheim, 1966, p. 124). Gombrich (1984/1992) later stated it as the most basic fact of aesthetic experience “that delight lies somewhere between boredom and confusion” (p. 9).

In later work, Arnheim refers to what he previously called order as orderliness, and describes the balance between orderliness and complexity as order. “A structure can be more or less orderly at any level of complexity. The level of ordered complexity is the level of order” (Arnheim, 1971, p. 51). Arnheim regarded ‘order’ (i.e., the balance between orderliness and complexity) as a necessary but not a sufficient condition for aesthetic appreciation (Arnheim, 1971).

Besides in the context of empirical aesthetics research, the balance between order and complexity has been discussed in many other research fields and application contexts, including acoustics (e.g., Fletcher, 2012), webpage design (e.g., Deng & Poole, 2012; Post, Nguyen, & Hekkert, 2017), environmental aesthetics (e.g., Nasar, 1994), fashion (e.g., Gray, Schmitt, Strohminger, & Kassam, 2014), and product design (e.g., Post et al., 2016). Post et al. (2016) define the principle of unity-in-variety as “the maximization of both unity and variety, in order to achieve a balance that offers the greatest aesthetic appreciation” (p. 142). They argue that unity (i.e., order) is the dominant factor in the relationship between unity and variety, and that unity (i.e., order) facilitates the appreciation of variety (i.e., complexity). They tested this idea of unity (i.e., order) and variety (i.e., complexity) being partial opposites while simultaneously contributing to aesthetic appreciation in the field of product design. They found that “the relationship between unity and variety is asymmetrical: the appreciation of variety is highly dependent on the presence of unity, whereas unity can be appreciated for its own sake” (Post et al., 2016, p. 150). From a study on webpage aesthetics, Post et al. (2017) concluded that “both unity and variety independently and positively influence aesthetic appreciation” (p. 48) and that “simultaneously maximizing unity and variety leads to an optimal balance where aesthetic appreciation is highest” (p. 48).

Various relationships between order and complexity. Although the principle of balancing order and complexity (i.e., unity in variety) has been perceived as one of the fundamental principles of aesthetics since ancient times, very few empirical studies have investigated the balance between order and complexity in a direct way (Post et al., 2016). In this project, we will follow Eysenck’s recommendations by trying to identify the elements determining order and complexity for this specific type of stimuli (i.e., images of neatly organized compositions) and trying to specify the formula relating both concepts when influencing aesthetic appreciation in an empirical way. However, before doing so, we will give our own view on the relation between order and complexity in determining aesthetic preference, based on the literature reviewed above.

We follow Arnheim (1966) in that the relationship between order and complexity is both antagonistic and complementary. On the one hand, order and complexity can be seen as partial opposites: Complexity reduces order, order reduces complexity (Arnheim, 1966). On the other hand, we expect order and complexity to complement each other. Whereas order needs complexity to show its structuring and clarifying potential, complexity needs order to be understood and appreciated.

Order and complexity as partial opposites. Although a lot of different types of order exist, we will use the example of similarity here. Similarity is a form of order (i.e., a way in which the information in a stimulus is ordered or organized), but it can also be seen as indicating a low level of complexity (i.e., a smaller amount and variety of information: simplicity). For example, noticing that all elements in the stimulus shown in Figure 1A are yellow cannot only make this image look less complex than other images, it is also a way to perceive the shown elements as a coherent whole. The similar color becomes a form of order or organization. This is one reason why we think order and complexity are interrelated and need to be studied together.

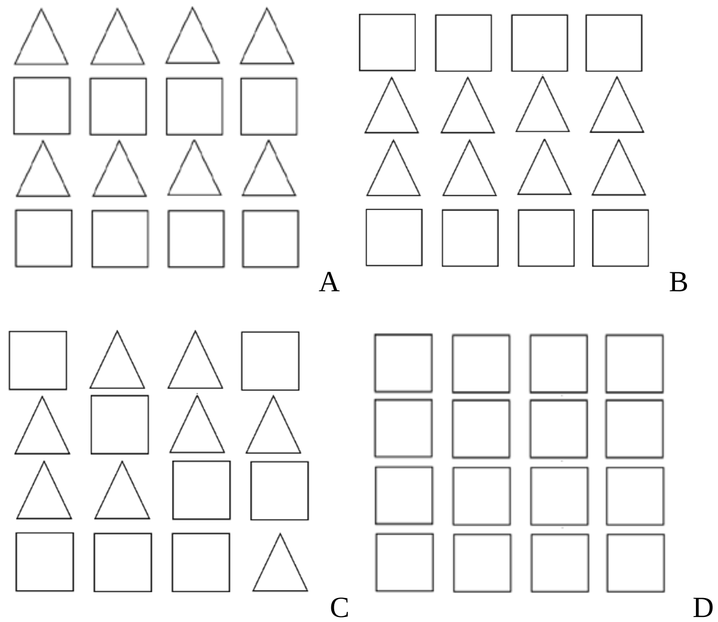


Figure 2. Illustration of order and complexity as partial complements. Order needs complexity to show its possibilities, complexity needs order to be appreciated.

Order and complexity as partial complements. A second reason is that order and complexity also need each other. On the one hand, order needs complexity to be able to show its possibilities. With more complexity (i.e., larger amount and variety of elements present in the stimulus), more possibilities arise for ordering or not ordering these elements. When complexity is low on a certain dimension, there are fewer possibilities to order the stimulus on that dimension than when complexity is high on that dimension (e.g., Figure 2A–C versus 2D). We expect the degree of complexity present in a stimulus to determine the order or disorder that can be present in that stimulus (both quantitatively and qualitatively). That is, stimuli with higher levels of complexity (a) can show a broader range of (dis)order levels than stimuli with low levels of complexity (e.g., Figure 2B and 2C); and (b) can show a broader range of qualitatively different orderings (without differences in the level of order that these ordering possibilities represent; e.g., alternation versus symmetry, see Figure 2A and 2B) than stimuli with lower levels of complexity (e.g., Figure 2D). Furthermore, ordered stimuli that lack complexity may be perceived as boring.

On the other hand, complexity also needs order. Stimuli that are highly complex (i.e., contain a large amount and variety of information) can only be ‘understood’ and aesthetically appreciated when the information is sufficiently structured. Completely unstructured or disordered stimuli may require too much energy or processing capacity and may be perceived as confusing. This is in line with the ideas of Post et al. (2016), who argued that unity (i.e., order) facilitates the appreciation of variety (i.e., complexity).

We summarize our ideas on the balance between order and complexity with the words of Moore (1917, p. 145):

> “Real art demands that we put forth a vigorous effort to comprehend a manifold of impressions in a single scheme of thought. When this effort is severe but successful, aesthetic joy is at its very pitch; when only a listless effort is called forth, the aesthetic work is trivial; when the effort demanded exceeds our powers of mental activity, or when there is violent thwarting of the powers called into play, we experience the painful shock of ugliness.”

In conclusion, it is hypothesized that order and complexity interact in how they influence aesthetic appreciation. While on the one hand contrasting each other, order and complexity also seem to depend on each other’s presence to optimize appreciation. However, individuals can vary in the amount of order they prefer relative to the amount of complexity that is present. In other words, individuals can vary in their preferred level of balance between order and complexity.

Alternative Accounts of Aesthetic Appreciation

Although we focus on the balance between order and complexity as a foundation for our research, other theories and principles of aesthetic appreciation have been proposed and used in empirical aesthetics too. Below we discuss some of the important theories of aesthetic appreciation and how they can be related to the balance between order and complexity.

Processing fluency and aesthetic appreciation. Reber and colleagues (e.g., Reber, Schwarz, & Winkielman, 2004) proposed processing fluency as a crucial determinant of aesthetic appreciation. In their fluency theory, they state that the more fluently a stimulus is processed, the more the stimulus will be preferred. Although it can explain many different phenomena in the field of experimental aesthetics (cf. Palmer et al., 2013), it would predict a linear decrease in preference with increasing complexity, as increased complexity is expected to require increased processing and complex stimuli would thus be processed less fluently than simpler ones. Several attempts have been made to reconcile preferences for complexity and the framework of processing fluency (e.g., Belke et al., 2015). Recently, Joye, Steg, Ünal, and Pals (2016) found evidence that some highly complex types of stimuli characterized by high internal repetition (e.g., fractals) can be processed fluently despite their visual complexity. More specifically, they asked participants to solve some puzzles while being exposed (or after being exposed) to either high-fractal or low-fractal stimuli. They found that participants solved the puzzles that were presented in combination with the high-fractal stimuli in less time and more accurately than puzzles presented in combination with the low-fractal stimuli. Furthermore, participants perceived the puzzles associated with the high-fractal stimuli as easier than the ones associated with the low-fractal stimuli.

By showing that processing fluency and complexity are not always negatively related, these findings are a first step in the reconciliation of the ideas and findings of fluency theory and theories predicting a positive relation between aesthetic appreciation and complexity. More specifically, the internal order present in a stimulus of high complexity (in the study of Joye and colleagues in the form of internal repetition) is expected to promote higher processing fluency and consequently also higher aesthetic appreciation. This is not only explicable in the framework of fluency theory, but also in the framework of a balance between order and complexity. More concretely, theories of the balance between order and complexity predict the high level of order (internal repetition) to compensate for high level of complexity in the high-fractal stimuli, leading to higher aesthetic appreciation.

Pleasure-Interest Model of Aesthetic Liking. The Pleasure-Interest Model of Aesthetic Liking (PIA model; Graf & Landwehr, 2015) distinguishes between pleasure-based and interest-based liking. It suggests a dual-process perspective on aesthetics, in which the first is stimulus-driven automatic processing and the second is perceiver-driven controlled processing. Immediate automatic processing is hypothesized to relate to pleasure, driven by a “gut-level fluency experience”, similar to the processing fluency perspective presented above. Perceiver-driven controlled processing is hypothesized to relate to evaluations of interest, triggered by a disfluency reduction. In a recent addition to the theory, Graf and Landwehr (2017) suggested that both pleasure and interest mediate the relationship between stimulus fluency and aesthetic liking.

When considering the relation between the PIA model and the principle of the balance between order and complexity, we could suggest that evaluations of interest (for rather complex, disfluent stimuli) only occur under perceiver-driven controlled processing because the ‘order’ resulting from the controlled processing (i.e., disfluency reduction) is necessary to counterbalance the complexity level of the stimulus. In contrast, evaluations of pleasure (for rather simple, fluent stimuli) do not ask for as much counterbalancing order, and thus proceed automatically and pleasure can be appreciated immediately.

More comprehensive theories on aesthetic appreciation. Also more comprehensive theories of aesthetic appreciation include order and complexity in their models. In the model of aesthetic appreciation and aesthetic judgments by Leder, Belke, Oeberst, and Augustin (2004), the principles of complexity and order are mentioned in the perceptual analysis of the stimulus (the first processing stage). Furthermore, the more conceptual and semantic types of order and complexity could play a role in later stages of the model (e.g., explicit classification, cognitive mastering). However, this is not explicitly mentioned by Leder et al. (2004).

Individual Differences in Aesthetic Appreciation, Order, and Complexity

As mentioned earlier, research on aesthetic appreciation can focus on both stimulus and person properties. Individual differences do not only play an important role in determining aesthetic appreciation as such (e.g., Güçlütürk et al., 2016; Jacobsen, 2004; Vessel & Rubin, 2010), but they may also affect the perception of order and complexity, and by that way have an effect on aesthetic appreciation. Furthermore, the type and strength of relationships between different aspects of order, complexity, and aesthetic appreciation may be subject to individual differences (e.g., Nadal, 2007).

Individual differences in aesthetic appreciation. Eysenck (1942) specified four types of individual difference factors influencing aesthetic appreciation. He proposed two general factors and two 'unique' factors (i.e., peculiar to each observer). The first general factor, the 'T'-factor (originally referring to good 'T'aste), was conceptualized as an innate factor of aesthetic appreciation (Eysenck, 1942). It has some resemblance to the general intelligence factor 'g', in the sense that individuals vary in their 'ability' for aesthetic appreciation. The second general factor, the 'K'-factor, distinguished people preferring simple and ordered stimuli from people preferring more complex and less ordered stimuli (Nadal, 2007). Eysenck (1942) related the 'K'-factor to the personality dimension of introversion-extraversion: "[…] the extravert prefers the simple, highly 'unified,' vividly coloured, modern type of picture; […] The introvert, on the other hand, prefers the complex, more diversified, less 'poster-colour' picture […]" (Eysenck, 1942, p. 353).

Regarding the two 'unique' factors (i.e., peculiar to each observer), Eysenck (1942) distinguished 'specific' and 'error' factors. Specific factors "include those various individual judgements which are based on private associations and experiences" (Eysenck, 1942, p. 354). Eysenck used the term 'error' factors to indicate those that show intraindividual variability.

In what follows, we will focus on individual differences related to (the balance between) order and complexity. Besides showing evidence that large individual differences in the aesthetic appreciation of order and/or complexity exist, we will discuss some factors that may explain these individual differences. However, we acknowledge that some of the factors mentioned under this section can also have an influence on aesthetic appreciation in general (e.g., Openness to Experience, art expertise).

Individual differences in the aesthetic appreciation of complexity. Large individual differences in the aesthetic appreciation of complexity have been found in several studies (e.g., Aitken, 1974; Güçlütürk et al., 2016; Jacobsen & Höfel, 2002; Tinio & Leder, 2009). When no individual differences were taken into account and only average relations were examined, Aitken (1974) found an inverted U-shaped relationship between complexity of random polygons and judgments of pleasingness. However, when looking at the individual participants, very diverse patterns emerged. In the same vein, Güçlütürk et al. (2016) argued that the inverted U-shaped function between complexity and liking is the result of averaging very different individual functions between complexity and liking. Whereas one group of participants showed a negative relationship between complexity ratings for several digitally generated grayscale images and liking, another group of participants demonstrated a positive relationship. Although these studies support the existence of individual differences in the aesthetic appreciation of complexity, they do not specify individual difference factors that could explain these diverging patterns. Below we discuss some personality and other individual difference variables that have been found to be possibly important in the aesthetic appreciation of complexity.

Perceptual ability and processing ability. Chevrier and Delorme (1980) related individual differences in the aesthetic appreciation of complexity to individual differences in perceptual ability. They reported that children scoring high on an overlapping figures test and an embedded figures test preferred stimuli of a higher level of complexity than low scoring children. Sherman, Grabowecky, and Suzuki (2015) found evidence that aesthetic appreciation is higher for artworks of a level of visual complexity that is compatible with the individual's working memory capacity. More specifically, they showed that individuals with lower visual working memory capacity tended to prefer artworks of a lower level of visual complexity than individuals with higher visual working memory capacity. Munsinger and Kessen (1964) reported that participants preferred an amount of cognitive uncertainty that matched their processing ability. Taken together, both differences in perceptual ability and processing ability seem to be related to aesthetic preferences for different levels of complexity.

Big Five personality traits. Besides individual differences in abilities, also individual differences in personality may relate to aesthetic preferences. Openness to Experience is not only related to a preference for art in general (Chamorro-Premuzic, Burke, Hsu, & Swami, 2010), it is also the Big Five trait that is most consistently found to relate to aesthetic preferences for complexity (e.g., in music: Rentfrow & Gosling, 2003; in works of art: Chamorro-Premuzic et al., 2010). Chamorro-Premuzic et al. (2010) additionally found a negative association between Conscientiousness and a preference for subjectively complex paintings (i.e., paintings that were perceived as containing many sophisticated and interconnected elements). Although Chamorro-Premuzic et al. (2010) predicted a positive association between Extraversion and preference for complexity and a negative association between Neuroticism and preference for complexity, their results did not support these last two hypotheses.

Need for cognitive structuring. A personality trait that seemed possibly more closely related to an individual's preferred level of complexity than the very general Big Five personality traits, is an individual's desire for closure. The need for cognitive closure is defined by Kruglanski (1990) as a desire for an answer, no matter which answer, in contrast to staying in a confused or ambiguous state. Cognitive closure allows individuals to predict future situations and to act based on these predictions. Related concepts are intolerance of ambiguity and intolerance of uncertainty. Intolerance of ambiguity refers to the tendency of an individual to interpret a present situation that is ambiguous (e.g., novel, complex, insoluble, unpredictable, uncertain) as a threat or a source of discomfort (Grenier, Barrette,

& Ladouceur, 2005). Previous studies reported a negative correlation of ambiguity intolerance with Openness to Experience (Bardi, Guerra, & Ramdeny, 2009) and with self-oriented perfectionism (Buhr & Dugas, 2006). Intolerance of uncertainty resembles the concept of intolerance of ambiguity but is more oriented toward future situations (Dugas et al., 2005). Uncertainty intolerance correlated positively with ambiguity intolerance, several aspects of the need for cognitive closure, Neuroticism, and Conscientiousness (Berenbaum, Bredemeier, & Thompson, 2008; Buhr & Dugas, 2006).

Cognitive structuring, described by Neuberg and Newsom (1993) as a way to reduce informational quantity and complexity, is one way to reach cognitive closure[4]. By creating and using abstract mental generalizations of previous experiences, individuals can understand the world without spending too much of their cognitive resources. The concept of a personal need for structure taps into the preference for structure and clarity as well as into reactions related to the absence of structure and clarity (Thompson, Naccarato, Parker, & Moskowitz, 2001). Neuberg and Newsom (1993) found that individuals scoring high on a measure of the personal need for structure tend to organize both social and non-social information in less complex ways than low scoring individuals. Furthermore, a person's need for structure related to a preference for predictability (Webster & Kruglanski, 1994) and to several Big Five factors. Neuberg and Newsom (1993) reported that a personal need for structure correlated negatively with Openness to Experience. Furthermore, the subaspect related to a desire and preference for structure was positively associated with Conscientiousness, whereas the subaspect related to negative reactions towards a lack of structure was positively associated with Neuroticism and negatively associated with Extraversion (Neuberg & Newsom, 1993).

Education level. When examining aesthetic preferences of students and manual workers, Francès (1976) found that students preferred complex stimuli over simple ones, whereas preferences for manual workers were reversed. When Francès (1976) looked at specific aspects determining aesthetic preferences, there were no differences between the groups for the preferred number of elements or the preferred level of heterogeneity of elements. However, differences between groups were significant for four other aspects of complexity. Francès (1976) argued that these differences could be explained based on the long-term novelty principle: Manual workers may not have had enough experience with the more complex types of stimuli to be able to aesthetically appreciate them in a positive way.

Expertise. "The more energetic and masterful the mind the more it will demand that the material of its art be really complex" (Moore, 1917, p. 135). Whereas Orr and Ohlsson (2005) did find relationships between complexity and aesthetic appreciation of jazz and bluegrass improvisations for non-experts, they did not find these relationships for expert musicians in the field. Munsinger and Kessen (1964) reported influences of experience on the relation between complexity and aesthetic

[4] Another way to reach cognitive closure that Neuberg and Newsom (1993) describe is avoiding encountering new information.

appreciation, both when resulting from specific professional training and when induced experimentally. In general, most studies found positive associations between art training and preferences for complexity (for a review, see Nadal, 2007). Furthermore, it has been suggested that art expertise may be positively associated with certain personality traits (e.g., Openness to Experience, need for cognition, and ambiguity tolerance; Belke, Leder, & Carbon, 2015).

Personality disorders. Some personality disorders have been related to the appreciation of complexity. King, Villeneuve, Post, Flowers, and Moonshine (1995) found positive associations between the narcissistic and borderline personality disorders and the aesthetic appreciation of complexity, whereas dependent personality disorder was negatively related to the appreciation of complexity.

Individual differences in the aesthetic appreciation of order. Although no studies have examined individual differences in the aesthetic appreciation of order directly, as far as we know, Palmer and Griscom (2013) examined the preference for harmonious stimuli (i.e., in the sense of being "good Gestalts") on four different dimensions (i.e., color, shape, spatial location, and music). They found strong correlations across the different domains. In other words, individuals with a high preference for harmony on the color dimension also tended to prefer more harmony on the other dimensions than individuals with a lower preference for harmony. Below we list some individual difference factors that could influence an individual's aesthetic appreciation of order.

Expertise. Specific training in a domain (i.e., spatial harmony for art training or musical harmony for music training) seemed to be related negatively to preference for harmony scores in that domain. Additionally, Palmer and Griscom (2013) argued that whether and how individual differences in the preference for harmony and complexity might be related could be an important aspect for further research.

Obsessive-compulsive tendencies. As obsessive-compulsive personality disorder (OCPD) is related to perfectionism and a preoccupation with details, order, and organization (American Psychiatric Association [APA], 2013 in Crego, Samuel, & Widiger, 2015) and compulsive ordering and arranging is one of the symptoms of obsessive-compulsive disorder (OCD; Radomsky & Rachman, 2004), it is hypothesized that obsessive-compulsive tendencies could be related to an aesthetic preference for order. Also Radomsky and Rachman (2004) argued that compulsive ordering and a drive for symmetry in OCD could be seen as extreme instances of a common preference for order and symmetry[5]. Obsessions and compulsions are known to occur in non-clinical populations (Gibbs, 1996). Radomsky and Rachman (2004) showed that the strength of the preference for pictures of orderly environments above pictures of disorderly environments correlated strongly positive with a measure of symmetry, ordering, and arranging behavior.

[5] Furthermore, they do not expect any differences in ordering and arranging behavior in OCD and OCPD, although it might have different functions in both disorders (Radomsky & Rachman, 2004).

In a review of research on OCD in non-clinical populations, Gibbs (1996) reported positive correlations between obsessive-compulsive symptoms and personality traits as neuroticism, introversion, and the need for control. Crego et al. (2015) found support for the statement that OCPD might be seen as a maladaptive variant of conscientiousness.

Individual differences in the perception of order and complexity. Individuals cannot only differ in how much they aesthetically appreciate a certain level of order, complexity, or the balance between order and complexity, they can also differ in their perception of order and complexity as such. This could amongst others be influenced by age, gender, or the definitions of order and complexity used by the individual.

Age. Chipman and Mendelson (1979) reported that sensitivity for different types of organization developed at different ages. They concluded that the type of organization present in a stimulus is important in understanding age-related differences in the perception of complexity. In other words, because young children do not pay attention yet to all types of organization, their perceived complexity scores may in some cases be determined more heavily by quantitative variables (rather than structural variables) than the scores of adults.

Gender. Nadal (2007) found that men and women differed in which aspects of complexity had the most influence on their perceived complexity ratings, at least for some types of stimuli (not for abstract decorative stimuli). Complexity scores given by men were best predicted by the number of elements in the stimuli and the variety of elements in the stimuli. The complexity scores given by women were best predicted by women's ratings of the number of elements in the stimuli and the variety of colours in the stimuli.

Definitions of order and complexity. Oliva et al. (2004) asked participants to hierarchically group images of real-world indoor scenes based on visual complexity. At each grouping stage participants had to describe the criteria they used for the categorization. When visual complexity was defined as related to how difficult it would be to remember the image and to give a verbal description of the image, participants reported criteria related to the number and variety of elements in the images. When visual complexity was defined as related to the structure of the scene image, participants gave more weight to dimensions related to order (i.e., clutter, symmetry, open space, and organization). Consequently, it could be that different individuals use different standard definitions of order and complexity, influencing their subjective ratings of the concepts (and possibly aesthetic appreciation).

Investigating the Balance Between Order and Complexity in the Aesthetic Preference for Neatly Organized Compositions

This project aimed to explore the variability in aesthetic appreciation of neatly organized compositions and which stimulus properties, person properties, and interactions between stimulus and person are associated with this variability. By extension, the study can also contribute to knowledge about human aesthetics and factors influencing human aesthetics in general. The focus was on both stimulus and person properties hypothesized to relate to (the balance between) order and complexity.

As discussed above (cf. “Aesthetics involves both stimulus and person”), earlier research indicated that pronounced individual differences in aesthetic preference can exist. This was also the case for real-world images when the semantic meaning of the images was de-emphasized (Vessel & Rubin, 2010). In the case of images of neatly organized compositions, the organization brought into these images could de-emphasize the semantic meaning of the presented objects. Therefore, we hypothesized that both stimulus and person properties as well as their interaction would be important in explaining aesthetic appreciation for images of neatly organized compositions.

To be able to take individual differences in aesthetic appreciation into account and to be able to investigate factors that could explain these differences, this project aimed for a rather large and very diverse sample of participants (i.e., in age, gender, and level of education). To reach this aim, we conducted the study online. Furthermore, both a Dutch and an English version of the study were administered to be able to investigate whether or not the results for the Dutch sample (mainly participants with the Belgian nationality) could be generalized to a more culturally diverse sample of English speaking participants[6].

Firstly, the project investigates the variability in aesthetic *preferences* for neatly organized compositions and how this variability can be explained. Participants’ aesthetic preferences were measured using a spatial 2AFC method. This means that participants indicated which one of two simultaneously presented images they preferred (cf. “Behavioral measures of aesthetic appreciation”). The disadvantage of the relative nature of the choice data collected did not outweigh the advantages of minimal memory load and easy response method, which is important for an online study. Also in the literature, this paradigm is described as the optimal task in most respects (Palmer et al., 2013).

As we expected that the principle of balancing order and complexity would be important in determining aesthetic preferences for neatly organized compositions, we selected stimulus and individual difference dimensions hypothesized to relate to (the balance between) order and complexity. These stimulus and person properties could then be related to the preferences for specific (types of) images. Because this was the first psychological study involving images of neatly organized compositions, the aim was to study a rather wide range of possibly relevant stimulus and individual difference dimensions in a rather exploratory fashion.

To explain variability in aesthetic preferences, we look at the relations of aesthetic preference with subjective order and complexity ratings as well as with objective indicators of complexity and individual differences in these relations. Additionally, we also investigate how aesthetic preferences relate to two other measures indicating positive aesthetic value, more specifically how soothing and fascinating each image was perceived to be, and whether individual differences exist in the relations of these soothingness and fascination ratings with aesthetic preferences.

[6] In a next phase of this project, also the data of a shortened Chinese version of this study will be analyzed to further investigate any cross-cultural similarities or differences in the aesthetic preferences for images of neatly organized compositions.

Secondly, the project will investigate which stimulus and person properties relate to *subjective order and complexity* ratings. By checking the relations with different stimulus properties, we could study which objective measures were most closely related to subjective measures of order and complexity. By checking individual differences in perceived order and complexity, we could investigate whether differences in preference are possibly a consequence of difference in perceived order or complexity of the images.

Thirdly, we will also investigate which stimulus and person properties relate to subjective judgments of *soothingness and fascination* of images of neatly organized compositions, as these different types of positive aesthetic value could interrelate with different levels of balance between order and complexity.

Exploring differences between stimuli. Although we agree with Berlyne (1960) that it is important to investigate both form and content (i.e., perceptual and conceptual) dimensions contributing to aesthetic appreciation, in the study reported below the main focus was on perceptual stimulus dimensions related to the balance between order and complexity. We focused on perceptual stimulus aspects as we thought that these perceptual factors would be relatively less influenced by individual differences that are very peculiar to specific stimuli. Further research could then focus on conceptual or semantic stimulus aspects involved in the aesthetic appreciation of neatly organized compositions.

Objective indicators of complexity that were computed were PHOG self-similarity, HOG complexity, HOG anisotropy, Fourier slope, and fractal dimension (cf. "Statistical image properties as objective measures of complexity"). These measures were chosen because of practical reasons (i.e., an acquainted professor could calculate these measures for us). Subjective order and complexity were measured in a second (optional) part of the study using rating scales from 1 (*not at all ordered/complex*) to 7 (*extremely ordered/complex*). Additionally, the stimulus pair selection for the 2AFC task and the coding scheme for the image pairs used in that task was based on our subjective judgments of differences in order and complexity on several perceptual stimulus dimensions, which could be seen as a rough subjective measure of differences in order and complexity. More information about the stimulus pair selection can be found in Appendix A and B.

In the second (optional) part of the study, we also asked participants to indicate how soothing and fascinating they experienced each image to be, to examine the function that images of neatly organized compositions could have for people in everyday life, and to explore which type of positive evaluation is most related to aesthetic preferences for neatly organized compositions. Do people appreciate an image more if it is soothing than when it is fascinating? Or do people also differ in their preferences for more soothing (i.e., probably more related to high order) or more fascinating (i.e., probably more related to high complexity) images? In lay interpretations, watching images of neatly organized compositions is often related to its soothing function. However, not everyone watches these images in his/her free time and individuals could differ in the reactions these pictures evoke.

Three short questions at the end of the first part of the study asked about (a) the overall pleasantness of the images they saw during the study; (b) whether they had seen similar images before; and (c) if they had seen similar images before, whether they had consciously sought them before. Participants indicating to have sought similar images before were asked to indicate why they had done so.

Exploring differences between individuals. Based on the lay interpretations that exist about the appreciation of neatly organized images and based on the reviewed literature about individual differences in aesthetic preferences presented above, we selected a number of individual difference dimensions related to (the balance between) order and complexity that, in interaction with the properties of the stimuli, could play a role in the aesthetic preference for certain neatly organized images. Each of these individual difference dimensions will be discussed below. Then questionnaires were selected to measure each of these dimensions in an accurate but concise way, to be able to measure different traits without asking too much time for participants to complete. We focused on individual differences related to personality, as we had clear hypotheses regarding the impact of these differences on aesthetic appreciation and because they seemed most relevant in a non-clinical sample of adult participants.

Big Five personality traits. As the Big Five personality traits are the individual difference traits most commonly investigated in current aesthetics research (Chamorro-Premuzic, Furnham, & Reimers, 2007), we included a measure of the Big Five personality traits in the study. Furthermore, investigating possible associations between very general personality dimensions and aesthetic preferences seemed a good start for this first study on images of neatly organized compositions. We selected the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991). As the BFI is freely available, widely used in online assessment (e.g., Srivastava, John, Gosling, & Potter, 2003), and recommended for use in cross-cultural settings (Schmitt, Allik, McCrae, & Benet-Martinez, 2007) as well as when a short Big Five instrument is needed (Denissen, Geenen, van Aken, Gosling, & Potter, 2008), the BFI was especially suited for this study.

Cognitive structuring. The Personal Need for Structure scale (PNS; Thompson et al., 1989, 1992) was selected because it directly measures one of the constructs of interest in a short and easy-to-administer way, and because of the previously found associations with a tendency to organize information in less complex ways (Neuberg & Newsom, 1993). It seemed more closely related to preferences for order than the broader concept measured by the Need For Closure Scale (Webster & Kruglanski, 1994). That is, the need for cognitive closure can be fulfilled in other ways besides increasing structure in the encountered information. An example of an alternative strategy is to avoid or limit the amount of information that is encountered (Neuberg & Newsom, 1993). Other alternative individual difference factors that were not included because they seemed very related to the PNS and/or because they seemed less suited for the purpose of this study were the Intolerance of

Uncertainty Scale (Freeston, Rhéaume, Letarte, Dugas, & Ladouceur, 1994) and the Ambiguity Tolerance–20 (Mac Donald, 1970).

Obsessive-compulsive tendencies. By including a measure related to obsessive-compulsive tendencies, and more specifically related to obsessive-compulsive symmetry, ordering, and arranging behavior, we aimed to investigate empirically the assumed associations of lay people between the appreciation of neatly organized compositions and obsessive-compulsive ordering behavior. The Symmetry, Ordering, and Arranging Questionnaire (SOAQ; Radomsky & Rachman, 2004) was selected as this questionnaire is focused on the ordering and arranging symptoms related to OCD, which was the main aspect of interest. Furthermore, the authors of the questionnaire argued that compulsive ordering and a drive for symmetry in OCD could be seen as extreme instances of a common preference for order and symmetry (Radomsky & Rachman, 2004; cf. "Individual differences in the aesthetic appreciation of complexity").

Research questions and hypotheses. Firstly, the variability in *aesthetic preferences for specific neatly organized compositions* is explored. We explore (a) which stimulus aspects relate to aesthetic preferences; and (b) whether there are individual differences in which stimulus aspects relate to aesthetic preference.

In general, we expected stimuli with the best balance between order and complexity to be preferred. This balance can be operationalized in several ways: as a sum, a difference, a multiplication, or a division. We did not have hypotheses about general preferences for order or complexity. However, we did expect preferences for order, preferences for complexity, and the optimal point of balance to be subject to individual differences.

We expected individuals scoring high on Openness to Experience to on average prefer more complex images and images with a higher level of complexity in comparison to the level of order than individuals scoring low on Openness to Experience. The same positive relation between complexity and aesthetic preferences – albeit to a lesser extent – was expected for individuals scoring high on Extraversion, individuals with a high level of education, and female participants. Additionally, the individuals scoring high on Openness or Extraversion may on average prefer less ordered images than low-scoring individuals.

To the contrary, we expected individuals scoring high on the PNS, Conscientiousness or Neuroticism to prefer less complex images and images with a lower level of complexity in comparison to the level of order than individuals scoring low on these personality dimensions. Additionally, high-scoring individuals may on average prefer more ordered images than low-scoring individuals. For individuals scoring high on the SOAQ, we in the first place expected a preference for more ordered images and images with a higher level of order in comparison to the level of complexity than scoring low on this measure. Additionally, high-scoring individuals may on average prefer less complex images than low-scoring individuals. We did not expect any pronounced influence of Agreeableness on individuals' preferences for order or complexity and did not have any hypotheses about the relation between age and preferences for order and complexity.

Although we did not have specific hypotheses about individual differences in preference for the most soothing or fascinating image in the pairs, we expected the ratings of how soothing an image was perceived to be to relate positively with subjective order and negatively with subjective complexity. For fascination ratings, we wanted to explore a possible interaction between order and complexity. More specifically, we expected that only images that are rated high on both order and complexity may be perceived as highly fascinating. Therefore, the individual differences in preference for order or complexity could consequently influence individual differences in preference for soothingness and fascination.

We also expected correlations between the different person properties. More specifically, we expected a positive correlation between the PNS and the SOAQ, and a negative correlation between Openness to Experience and both the PNS and the SOAQ. To a lesser extent, we also expected negative correlations between Extraversion and both PNS and SOAQ, as well as positive correlations of PNS and SOAQ with Conscientiousness and Neuroticism.

Secondly, the variability in *order and complexity ratings* is explored. We explore (a) which stimulus properties are associated with subjective order and complexity of an image; (b) whether individuals differ in their average ratings of order and complexity (across all images) and which person properties can explain these overall differences; and (c) whether individuals differ in which stimulus properties are associated with subjective order and complexity and which person properties can explain these differences.

We expected subjective ratings for complexity to relate positively with HOG complexity, self-similarity, fractal dimension, and Fourier slope, and negatively with anisotropy. If any influence of the person properties would exist, we expected individuals scoring high on the PNS to perceive the images on average as more complex (and possibly less ordered) than low-scoring individuals. Similar but probably smaller effects could exist for individuals scoring high on Conscientiousness and Neuroticism. Reversed effects could exist for individuals scoring high on Openness to Experience and Extraversion. Additionally, we expected individuals scoring high on the SOAQ to perceive the images as less ordered (and possibly more complex) than individuals with low SOAQ scores. However, as there is very little research about individual differences in perceived order or complexity, these hypotheses should be viewed with the necessary caution. We did not expect influences of gender, age, or level of education on the ratings of order and complexity. If an effect of these variables would exist, we would expect that highly-educated and female individuals would perceive the images as less complex than individuals with a lower level of education and male individuals respectively.

Although not central to our research questions, we did expect to find correlations between the different statistical image properties similar to those found earlier studies for different types of stimuli (e.g., Braun et al., 2013). More concretely, we expected a strong positive correlation between HOG complexity and fractal dimension, and possibly positive correlations between HOG complexity, fractal dimension, self-similarity, and Fourier slope. Anisotropy could relate negatively to the other statistical measures. However, the correlations could be different from previously reported ones because we are using a different type of images.

Thirdly, the variability in *soothingness and fascination ratings* is explored. We explore (a) which stimulus properties are associated with how soothing or fascinating an image is perceived to be; (b) whether individuals differ in their average soothingness and fascination ratings (across all images) and which person properties can explain these overall differences; and (c) whether individuals differ in which stimulus properties are associated with how soothing or fascinating an image is perceived to be and which person properties can explain these differences.

As mentioned earlier in this section, we expected the ratings of how soothing an image was perceived to be to relate positively with subjective order and negatively with subjective complexity. For fascination ratings, we wanted to explore a possible interaction between order and complexity. More specifically, we expected that only images that are rated high on both order and complexity may be perceived as highly fascinating. We did not have any specific hypotheses about the relation of soothingness and fascination with objective measures of complexity. However, as the stimulus properties would show the expected relations with ratings of order and complexity, they could consequently also be related to soothingness and fascination ratings.

We did not have specific hypotheses about the relations between soothingness and fascination and different person properties, but if the person properties would show the expected relations with subjective order and complexity, and subjective order and complexity ratings would relate to ratings of soothingness and fascination, relations between person properties and the ratings of soothingness and fascination would follow as a consequence.

Methods

Participants

Anyone with an age between 16 and 100 years old and able to understand Dutch or English instructions could participate. There were no restrictions regarding nationality or mother tongue. Participants were recruited via personal contacts of the researchers, social media, and offline advertisements in public places and university buildings. Participation was completely voluntarily: No monetary reward was offered for participation.

The first part of the study was completed by 421 participants between the ages of 16 and 77 years (274 women, 147 men, M_{age} = 39.8 years, SD_{age} = 15.8 years)[7]. In the analyses, the data of 415 participants between the ages of 16 and 77 were used (273 women, 142 men, M_{age} = 39.7 years, SD_{age} = 15.8 years), of which 359 completed the Dutch and 56 the English version of the study. The answers

[7] In total, the first part of the study was started 486 times. Sixty-five participations to the first part were not completed.

of 5 participants were removed because their preference data showed increasingly less variability in choosing the left versus the right image to the end of the 2AFC task. The answers of one additional participant were removed because the preference data showed no variability (i.e., the participant always indicated the right image). Of the 421 participants to the first part, 84 participants between the ages of 20 and 75 years completed the second (optional) part of the study (56 women, 28 men, M_{age} = 43.4 years, SD_{age} = 16.7 years)[8]. Only 8 participants to the second part completed the English version, whereas 76 completed the Dutch version. Further details on the demographic characteristics of the sample are shown in Supplementary Tables S1 and S2 in Appendix C. The study received ethical approval from the Social and Societal Ethics Committee of KU Leuven (dossier no. G-2016 04 547).

Materials

Image data. For the purpose of this study, 100 image pairs involving 184 different images—16 images were included twice—were selected. The attention to ecological validity and the wish to explore several dimensions possibly involved in the preferences for these images made it hard to obtain an equal number of image pairs for each stimulus dimension involved. Therefore, it was chosen to refrain from an equal number of pairs per dimension. In a first stage, images involving neatly organized compositions were collected from online sources (blogs, websites, etc.), constituting an extensive database including more than 1,000 images. Secondly, a selection was made from this database: As explained earlier (cf. 'Examining factors related to order and complexity'), images containing important aspects of order and complexity on a more conceptual level and only few perceptual aspects of order or complexity were not included. Thirdly, from the remaining images, it was attempted to construct pairs that only differed in order and/or complexity on one or two of the dimensions included in the study (i.e., color, texture, configuration, number of objects, type of objects, and perspective), while order and complexity on other dimensions (both those included in the study and others that were not included, e.g., background) were kept as constant as possible (for more information on the stimulus pair selection and the perceptual stimulus dimensions included, see Appendix A). As this study tried to find a balance between ecological validity and experimental control, only existing images were used, but they were sometimes slightly adapted to better fit the purpose of the study (i.e., to make the images within a pair more comparable on dimensions that were not meant to differ within that image pair). These adaptations included using only a subsection of the existing image, extrapolating it by copying a row, or trimming the borders of the image. Image pairs were constructed to differ in order and/or complexity on one of the following stimulus dimensions: color (25 pairs), texture (39 pairs), configuration (10 pairs), number of objects (10 pairs), perspective (3 pairs), and type of objects (i.e. same or different; 2 pairs). Examples of these pair types can be found in the Figures A1–A5 in Appendix A. Within these pair types, not all pairs tested the same

[8] In total, the second part of the study was started 123 times. Thirty-nine participations to the second part were not completed.

difference. For example, in the color pairs, a distinction can be made between pairs testing low versus medium color complexity and pairs testing medium versus high color complexity (see Figures A1.a and A1.b for an example). Eleven pairs were hypothesized to include important differences on two dimensions. Factor combinations included were: color and configuration (2 pairs); color and type (1 pair); color and texture (5 pairs); and texture and number of objects (3 pairs). Example stimulus pairs of these factor combinations can be found in Figure A5 in Appendix A. A classification of the image pairs was made, indicating for each image pair on which dimensions the images differed in order and/or complexity, and indicating which image was supposed to be more complex and/or less ordered (see Appendix B). It must be noted that most of the stimulus pairs included in the study were assumed to vary in complexity or in both complexity and order rather than in order only. However, the average order and complexity ratings for the images (of which the collection will be discussed later in this Method section) can give an indication of differences between the images in perceived order (although those ratings were general and thus not for each stimulus dimension separately).

Big Five Inventory. The Big Five Inventory (BFI; John et al., 1991) is a short self-report instrument (44 items) to measure the Big Five dimensions of personality (i.e., Extraversion, Neuroticism, Conscientiousness, Agreeableness, and Openness to Experience; McCrae & John, 1992) when no further differentiation between individual facets is needed (John, Naumann, & Soto, 2008). This measure uses short phrases instead of single adjectives as it was found that such items are answered more consistently when they are accompanied by definitions or elaborations (Goldberg & Kilkowski, 1985). Each BFI scale includes between eight and ten items and participants have to indicate their agreement with each statement using a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). In U.S. and Canadian samples, the English version of the BFI scales showed substantial convergent and divergent validity with other Big Five measures as well as with peer ratings. Reliability of the scales ranged from .75 to .90 (John et al., 2008). Also the Dutch version of the BFI (created and validated by Denissen et al., 2008) showed good psychometric properties. It was found to be factorially equivalent to the English original and the relative independence and internal consistency of the scales was preserved[9].

Personal Need for Structure scale. The Personal Need for Structure scale (PNS; Thompson et al., 2001) is a self-report measure of 12 items created to measure several aspects of the desire for simple structure. It is commonly calculated using a single factor that captures participants’ overall tendency to prefer simple structure (Thompson et al., 2001), but other authors suggest to use a two-factor interpretation in which item 5 (“I enjoy being spontaneous.”) is excluded and the scale captures both the Desire for Structure (DFS) and the Response to Lack of Structure (RLS; Cavazos, Judice-

[9] As in Denissen et al. (2008), a principal components analysis with varimax rotation of the 44 BFI items was conducted on the data of participants to the Dutch version of the current study (*N* = 359). The analysis yielded similar results to those of Denissen et al. (2008). The five factor solution explained 44% of the variance, with absolute primary loadings ranging from .34 to .77. However, some absolute cross-loadings went up to .50.

Campbell, & Ditzfeld, 2012; Neuberg & Newsom, 1993). Although those two subscales are found to correlate quite highly (rs = .54–.75; Neuberg & Newsom, 1993), it might be worth to look at the subscale scores as well, as differential relations were found between the subscales and dimensions of the Big Five (Neuberg & Newsom, 1993). Participants were asked to respond using a 6-point rating scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*).

Several versions of the PNS scale with slightly different item wordings and item orders circulate. The version that was used in this study is the version presented by Thompson et al. (1989), because the Dutch translation that was used (Het Nieuwe TeamWerken, 2015) was based on this version of the scale. Although no validation information was available about the Dutch translation of the PNS scale, the English version was found to have sufficient reliability and convergent and discriminant validity (Neuberg & Newsom, 1993; Thompson et al., 2001)[10].

Symmetry, Ordering, and Arranging Questionnaire. The Symmetry, Ordering, and Arranging Questionnaire (SOAQ; Radomsky & Rachman, 2004) is a self-report measure developed to assess beliefs and behavior associated with compulsive ordering and arranging, features of obsessive-compulsive disorder (OCD). Participants are asked to rate how strongly they agree with each statement on a 5-point scale ranging from 0 (*not at all*) to 4 (extremely). The authors suggest that compulsive ordering and a drive for symmetry are extreme manifestations of the common preference for order and symmetry (Radomsky & Rachman, 2004). Based on a sample of undergraduate students, the English SOAQ has a very good inter-item and test-retest reliability as well as very good convergent and divergent validity (Radomsky & Rachman, 2004). As the SOAQ was translated to Dutch for the purpose of this study, no validation information for the Dutch version was available[11].

Image questionnaire. Three additional questions were asked to participants regarding their earlier experiences with the type of images involved in this study, as well as regarding their liking of the presented images. More specifically, participants were asked about the pleasantness of looking at the images (i.e., “How pleasant did you find the images to look at in general?”). They could answer on a scale ranging from 1 (*not at all pleasant*) to 7 (*very pleasant*). Secondly, participants were asked to indicate their previous experience with this type of images (i.e., “Did you see similar images before?”) on a scale ranging from 0 (*never*) to 4 (*very often*). If they indicated to have seen similar images

[10] Two principal components analyses with varimax rotation of the 12 PNS items were conducted on the data of participants to the Dutch version of the current study (*N* = 359), one exploring a one-factor solution and one exploring a two-factor solution. The one-factor solution explained 38% of the variance, with loadings ranging from .40 to .77. As one item (i.e., “I enjoy being spontaneous”) showed a somewhat lower loading (i.e., .40) on the common factor than the other items (i.e., ranging from .51 to .77), and this item was sometimes dropped from the scale in earlier studies (e.g., Neuberg & Newsom, 1993), we decided to remove this item from the scale in this study as well. The two-factor solution did not resemble the two factors of Neuberg and Newsom (1993). Therefore, in the analyses using the PNS, the one-factor solution including 11 items was used.

[11] A principal components analysis with varimax rotation of the 20 SOAQ items was conducted on the data of participants to the Dutch version (*N* = 359) to test the one-factor solution. The one-factor solution explained 57% of the variance, with loadings ranging from .60 to .84.

before, they were additionally asked to indicate whether they had already consciously sought for similar images (i.e., “If so, did you already consciously search for similar images?”) on the same scale ranging from 0 (*never*) to 4 (*very often*). When participants indicated that they had already sought for similar images, they were asked to describe why they had done so (i.e., “If so, why?”).

Procedure

Data were collected online from May 2016 until January 2017. When participants visited the webpage of the questionnaire, they were provided with a short description of the study and were asked for their informed consent. If participants agreed to participate, they were asked to log in based on the provision of an e-mail address, to complete some basic demographic information (i.e., gender, age, mother tongue, and highest education level), and to complete the Big Five Inventory (BFI, 44 items).

Then, a 2AFC image task was conducted, in which participants had to indicate which of two simultaneously shown stimuli (i.e., images) they preferred. Participants were presented with 100 image pairs and were asked each time to click on the image they preferred. The image pairs were presented in a semi-random order. It was made sure that image pairs including images that were used in more than one pair were not shown immediately after each other. Additionally, the position (i.e., left or right) of the images hypothesized to be the most complex was counterbalanced between participants. When a participant had indicated his or her preference for one of the images, the test automatically continued to the next pair.

Afterwards, participants completed the Personal Need for Structure scale (PNS, 12 items) and the Symmetry, Ordering, and Arranging Questionnaire (SOAQ, 20 items), as well as the short questionnaire about their previous experience with and aesthetic appreciation of the type of images that were shown. After completing the questionnaires, participants were provided with a short debriefing text and had the possibility to indicate if they wanted to participate in an additional task at a later moment and if they wanted to be informed about the results of the study. This first part of the study was expected to take approximately half an hour[12].

If participants indicated that they were willing to participate in an additional task at a later moment, an automatically generated e-mail was sent to invite them for participation. As this optional task was expected to take quite a long time and to be quite exhausting, participants were asked in this e-mail to take enough time for participation and to not lose concentration if they decided to participate. When participants visited the webpage of this additional task, they were asked for an additional informed consent, logged in with the same e-mail address they had provided in the first part of the study, and were shown six example images (not part of the test images; see Figure 1) to remind them of the type of images they would see. The latter was done to remind participants of the variety of the images in the set, and to diminish context and order effects on the ratings (cf. “Behavioral measures of

[12] In reality, the mean completion time for the first part of the study was 24 minutes and 4 seconds. The median completion time was 16 minutes and 47 seconds.

aesthetic appreciation”). Then, the same images as those in the 2AFC image task were presented (184 images) in a randomized order. This time, participants were asked to rate each image separately on four characteristics (i.e. ordered, complex, soothing, and fascinating) using a 7-point scale ranging from 1 (*not at all*) to 7 (*extremely*). The order of the four characteristics was randomized over participants but stayed constant for each individual participant. This additional task was expected to take approximately 45 minutes, but participation durations showed that this might have been an underestimation of the mean time needed to complete the task[13]. After rating all 184 images, a short debriefing text was presented and participants were thanked for their participation. Additionally, participants could indicate if they wanted to be informed about the results of the study and if they were willing to participate in further research concerning the topic.

Image Calculations

For each of the 184 images used in the study, five statistical image properties were calculated: self-similarity, complexity, anisotropy, Fourier slope, and fractal dimension. The objective measures of self-similarity, complexity, and anisotropy were calculated using the Pyramid of Histograms of Orientation Gradients (PHOG) method (Bosch, Tisserman, & Munoz, 2007). The PHOG method involves three steps to obtain a gradient image for a color image. First, the original color image is divided into three different channels: a luminance channel, a red-green opponent channel, and a blue-yellow opponent channel. Second, the gradient image is calculated for each of the three channels. This gradient image indicates for each pixel the magnitude of change in that channel (for a given direction). Third, the maximum value for each pixel is selected from the three gradient images. The image resulting from the selected maximum value per pixel is called the gradient image (for a more elaborate description, see Appendix A from Braun et al., 2013). To calculate the PHOG-based measures, all images were rescaled to 100,000 pixels total size to make the values obtained comparable between images.

After calculating the gradient image, the Histograms of Oriented Gradient (HOG) features of this gradient image are computed (Dalal & Triggs, 2005). Starting from the gradient image, the orientations of the gradients are divided into a specified number of bins (in which the value for the number of bins and the range of orientations used in the calculation are arbitrary). In the calculations for this project, 16 bins covering 360 degrees of orientations were used. The strength of all gradients is then calculated for each bin of orientations. Finally, the histogram values are normalized, to make sure that the strength values for all bins of orientations sum to one. These calculated histogram values are then used to compute the measures of self-similarity, complexity, and anisotropy (as described below).

[13] In reality, the mean completion time for the second part of the study was 1 hour, 20 minutes, and 3 seconds. The median completion time was 57 minutes and 52 seconds. It has to be noted that some participants took long breaks, which influenced completion times.

For the Fourier analysis, the images were rescaled to 512 x 512 pixels (after padding with gray values to obtain square images). The range of frequencies analyzed was 5–256 cycles per image to avoid artifacts at the extremes of the spectrum.

PHOG self-similarity. The calculation of the self-similarity measure involves four steps. First, the HOG feature for the entire gradient image (level 0) is calculated. Second, the image is divided in four equal parts and HOG features for each part are calculated separately (level 1). Third, each part described in the second step is again divided in four equal parts and HOG features for each of the 16 parts are calculated separately (level 2). Fourth, each part described in the third step is again divided in four equal parts and HOG features are calculated separately for each of the 64 parts of the image (level 3). The PHOG representation of an image includes all the calculated levels (in this case, level 0–3). The Histogram Intersection Kernel (HIK; Barla, Franceschi, Odone, & Verri, 2002) is used to determine how similar two HOG features are. In this project, the ground approach was used to calculate self-similarity: the self-similarity for an image at any level was calculated by taking the median value of similarity between the HOG features for the subsections at a certain level and the HOG feature at the ground level (level 0). For the final self-similarity value, the self-similarity values computed for levels 1–3 were averaged, with the same weight given to each level. A high value for PHOG self-similarity indicates that the parts of the image are very similar to the image as a whole (Lyssenko et al., 2016)**.**

HOG complexity. Objective complexity using the HOG method is calculated by taking the mean of all gradient strengths in the gradient image. An image has high HOG-based complexity when there are large changes in luminance or color in the image, whereas an image has low HOG-based complexity when there are only small changes in luminance or color in the image.

HOG anisotropy. Anisotropy is calculated by taking the standard deviation over the HOG feature values (gradient strengths) for the image at the last level (in this case level 3). A high value for HOG anisotropy indicates that the magnitude of the changes in luminance or color is higher for some orientations than for other orientations in the image, whereas a low value indicates that the magnitude of the changes in luminance or color is similar across orientations.

Fourier slope. Fourier slope, or more specifically the slope of log-log plots of radially averaged Fourier power spectrum, was calculated as described in Redies, Hänisch, Blickhan, and Denzler (2007). As indicated earlier (cf. “Statistical image properties as objective measures of complexity”), in images with a Fourier slope higher than -2 the high spatial frequencies (representing fine detail) are relatively more prominent than the low spatial frequencies (representing coarse detail), images with a slope of -2 have scale-invariant properties, and in images with a slope lower than -2 the low spatial frequencies are relatively more prominent. Also the Fourier sigma was calculated for each image (for the calculation procedure, see Redies, Hänisch, et al., 2007). The Fourier sigma indicates how good the fit of the Fourier slope is and how useful it is to determine the Fourier slope of the plot.

Fractal dimension. Fractal dimension was computed with the box-counting method, as described in Braun et al. (2013). When using the box-counting method (see also Hagerhall et al., 2004), the image is divided in squares or boxes of equal size. Then the proportion of squares occupied by the pattern is calculated as well as the proportion of squares that is empty. The process is repeated for squares of increasingly small sizes.

Data Analysis

Reasons for data exclusion were (a) not having completed the part of the study in question; (b) showing increasingly less variability in choosing the left versus the right image to the end of the 2AFC task; and (c) showing no variability at all in choosing the left versus the right image in the 2AFC task (i.e., one participant always indicated the right image).

As the a priori coding of the image pairs (i.e., attempted to code the more complex image in the pair as ID1, the less complex image as ID0) was rather arbitrarily defined, we created new coding schemes based on the empirical data. More specifically, a preference coding (ID1 is preferred over ID0 by most participants), an order coding (ID1 is on average perceived as less ordered than ID0), a complexity coding (ID1 is on average perceived as more complex than ID0), a soothingness (ID1 is on average perceived as more soothing than ID0), and a fascination coding (ID1 is on average perceived as more fascinating than ID0) were determined.

For the ratings of order, complexity, soothingness, and fascination, average scores per image (N = 184) and per participant (N = 84) were calculated. For the average ratings per image, first the standardized score for each image was calculated per participant per rating scale, to eliminate biases in the use of the rating scale. Then, the average rating per image per rating scale (across participants) was calculated. Finally, these average ratings were again standardized. Also the statistical image properties were standardized. To calculate the average ratings per participant, the ratings were not first standardized per participant, as we were also interested in explaining different uses of the scale (e.g., higher scores on average could indicate that a person experiences the images overall as more ordered, complex, soothing, or fascinating). We did, however, standardize the scores on the individual difference variables to avoid problems related to collinearity of the predictors.

To explore the variability in aesthetic preferences for neatly organized compositions, we calculated the proportion of participants that preferred a specific image in each of the image pairs using all these different codings. We also calculated the proportion of image pairs in which each participant preferred the more ordered, complex, soothing, or fascinating image in the pairs.

To explore the structure in the preferences for certain images based on both image pair and person characteristics simultaneously, the Hierarchical Classes Approach (De Boeck, Rosenberg, & Van Mechelen, 1993; Rosenberg, Van Mechelen, & De Boeck, 1996) was used. With this technique, participants were clustered in classes of participants based on their aesthetic preferences. Also the image pairs were clustered in classes based on the overlap in participants preferring the same image in those pairs. Furthermore, the analysis results in a hierarchy of item pairs and participant classes, in which item pairs or participants who have more 1-scores are presented higher in the structure. Subsequently, the goal was to explore whether and how the structure of the image pairs and participants could be related to the different stimulus dimensions and individual differences hypothesized to play a role in the preferences for certain images.

To explore which stimulus aspects are associated with aesthetic preference, we calculated the mean scores for the order and complexity measures for on average preferred and on average non-preferred images separately. Furthermore, we calculated Pearson product-moment correlations between the different image pair measures and proportions of preference for the on average preferred image in the pairs. We also looked into the number of pairs in which the most ordered, complex, soothing, or fascinating image was significantly preferred above the other image in the pair.

To explore individual differences in preference for order, complexity, soothingness, or fascination, Pearson correlation coefficients between the different preferences and the measured person properties were calculated. Afterwards, multiple linear regression analyses were conducted to investigate which person properties could significantly predict the individual differences in preference.

To explore the variability in order and complexity ratings between images, we computed Pearson correlation coefficients between the average order and complexity ratings per image and the other image measures available. Additionally, multiple regression analyses were conducted to look into which image measures could best predict the average ratings for order and complexity. The same procedure was followed for exploring variability in soothingness and fascination ratings between images.

To explore the variability in order and complexity ratings between individuals, we computed Pearson correlation coefficients between the average order and complexity ratings per individual and the other individual difference measures available. When multiple significant correlations were found, multiple regression analyses were conducted to look into which individual difference measures could best predict the average ratings for order and complexity. The same procedure was followed for exploring variability in soothingness and fascination ratings between individuals.

Except for the hierarchical classes analysis, all data processing and analyses were conducted using the statistical program R (R Core Team, 2017).

Results

After describing the overall pleasantness of and participants' previous experience with the images, we first discuss the variability in aesthetic preferences for neatly organized compositions, which stimulus properties are associated with aesthetic preferences in general, and which person properties interact with these stimulus properties to predict individual differences in aesthetic appreciation. Secondly, we look into the variability in judgments of order and complexity, and stimulus and person properties predicting average and individual judgments of order and complexity. Thirdly, the variability in ratings of soothingness and fascination is explored, as well as stimulus and person properties predicting average as well as individual ratings of soothingness and fascination.

Overall Pleasantness of and Experience with Neatly Organized Compositions

Overall, most participants perceived the images as pleasant to look at (see Figure 3). Most participants did not regularly see similar images before (see Figure 4A). Of the participants who had seen similar images before, almost no participants indicated to search for this type of images regularly (see Figure 4B).

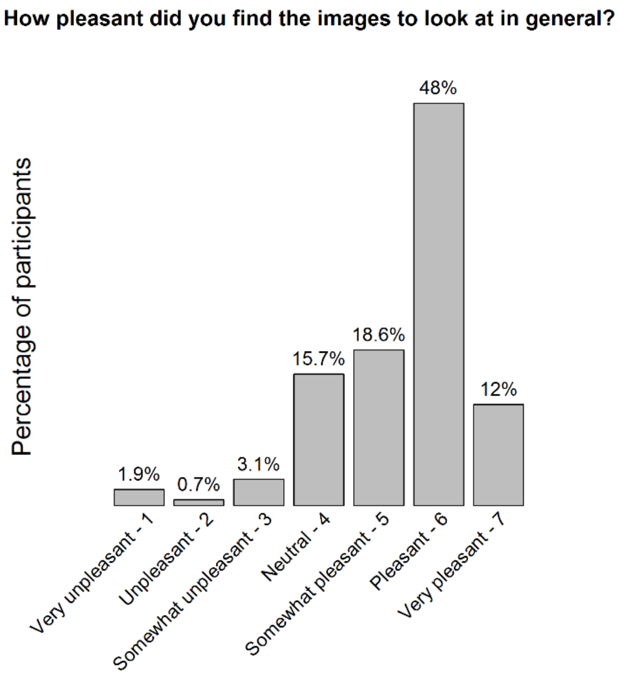


Figure 3. Distribution of the overall pleasantness score participants gave to the images.

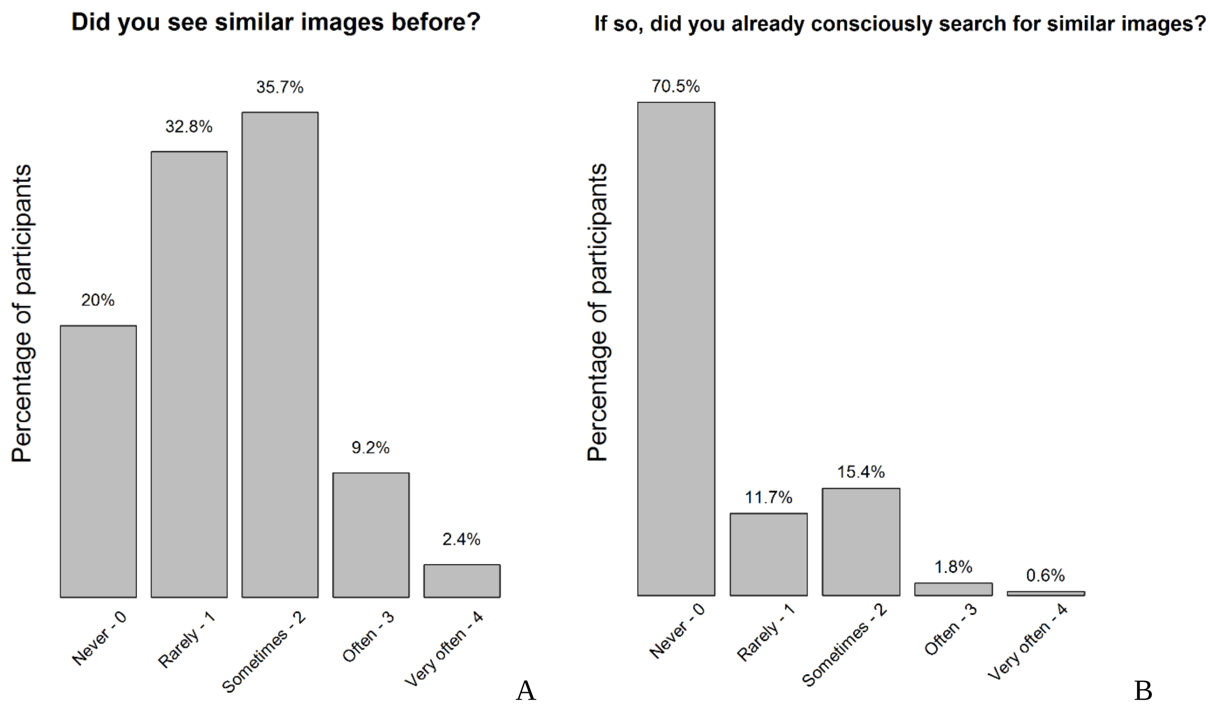


Figure 4. Distribution of how often participants saw (A) and searched for (B) similar images before.

Aesthetic Preferences for Neatly Organized Compositions

Amount of variability between image pairs. The amount of variability in the proportion of participants that preferred a certain image in the pairs is dependent on the coding of the image pairs that is used (cf. "Data Analysis"). As expected, image pairs differed in the amount of individual variation in preference that was present: Proportions for the preferred image varied between 0.50 and 0.81 (*M* = .6092, *SD* = .0724). As all preference proportions differed from 0 or 1, there were individual differences in aesthetic preferences: Individuals differed in which image within the pair they preferred. However, in 60 of the 100 image pairs tested, one of the images was preferred significantly more often than would be expected by chance (proportions between 0.59 and 0.81), based on exact binomial tests, $p < .001$ (two-tailed). This indicates that also some general preferences exist for certain images within the pairs for a large part of the pairs. The histograms of the proportion of participants that preferred the most ordered, most complex, most soothing, most fascinating, and (on average) most preferred image in the pairs are included as Supplementary Figure S1 in Appendix D.

Amount of variability between individuals. Also the variability between individuals is dependent on the image pair coding used (cf. "Data Analysis"). Preferences for the most ordered image in the pair varied between individuals, with preference proportions ranging from .32 to .74 (*M* = .5279, *SD* = .0740). Preferences for the most complex image in the pair ranged from .22 to .74 (*M* = .5045, *SD* = .0981). Preferences for the most soothing and the most fascinating image in the pair ranged from .36 to .74 (*M* = .5506, *SD* = .0657) and from .38 to .77 (*M* = .5767, *SD* = .0772) respectively. Preferences for the image in the pair that was preferred by most participants ranged from .44 to .77 (*M* = .6092, *SD* = .0669). The histograms of the proportion of image pairs in which the most ordered, complex, soothing, fascinating, or the on average preferred image was preferred across participants are included as Supplementary Figure S2 in Appendix D.

Effect of image position. Averaged over all image pairs, which image within each pair was positioned left or right (i.e., id0 or id1) had no influence on the proportion of preference for that image, based on a two-sample test for equality of proportions, $\chi^2(1) = 2.45$, $p = .12$.

Clustering of participants and image pairs based on aesthetic preferences. We performed a disjunctive hierarchical classes analyses using the HICLAS software (Version 0.9 beta; Ceulemans & Meers, 2005) on the preference data to identify clusters of participants and image pairs. The data consisted of a 415 participants x 100 image pairs matrix. The entries in the analyzed matrix were the aesthetic preferences for each participant per image pair. The analysis was repeated for different codings: (a) 1 indicating that the most ordered image was chosen; (b) 1 indicating that the least ordered image was chosen; (c) 1 indicating that the most complex image was chosen; and (d) 1 indicating that the least complex image was chosen. For each of these recoded data matrices, HICLAS solutions including 1-5 bundles were obtained.

However, as the goodness-of-fit values for all the fitted models were rather low (below .56), and the proportion of discrepancies between the models and the data were rather high (above .31), we decided to not report the HICLAS solutions and subsequent analyses based on these solutions.

As the analyses based on the HICLAS solutions could not be conducted in the planned way, the preference data and its relation to person and image pair properties were explored in other ways.

Relations between aesthetic preferences per image pair and image pair measures. Regardless of the size of the difference, in 78 of the 100 image pairs the most fascinating image was chosen. In 68 cases the most soothing image was chosen, in 60 cases the most ordered image, and in 54 cases the most complex image in the pair.

For the 60 image pairs where there was a clear preference for one of the pairs, in 49 cases this was the most fascinating image, in 47 cases the most soothing, in 38 cases the most ordered image, and in 31 cases the most complex image.

Differences between on average preferred and non-preferred images. Paired samples t-tests were conducted to explore whether the images that were significantly preferred differed from the non-preferred images in one or more of the image measures. Across the 60 image pairs in which there was a significant preference for one of the images in the pair, the preferred images were on average slightly more ordered (M = 0.2634, SD = 0.9206) than the non-preferred images (M = – 0.1546, SD = 1.0852), $t(59)$ = – 2.9463, p = .005. There was also a significant difference in soothingness, $t(59)$ = – 6.2209, p < .0001, and fascination, $t(59)$ = – 7.1088, p < .0001, between the preferred and non-preferred images. The preferred images were on average more soothing (M = 0.3325, SD = 1.0407) and more fascinating (M = 0.2664, SD = 1.0158) than the non-preferred images (M = – 0.1524, SD = 0.9205 and M = – 0.2562, SD = 0.9011, respectively). The standardized Fourier slope of preferred images (M = 0.13516, SD = 0.8471) was significantly higher than the standardized Fourier slope of non-preferred images (M = – 0.2258, SD = 0.8918), $t(59)$ = – 2.8324, p = .006. The differences in subjective complexity, self-similarity, HOG complexity, anisotropy, and fractal dimension between the preferred and non-preferred images were not significant at the α = .01–level, although the differences in self-similarity, anisotropy, and fractal dimension approached significance (ps < .05). We also explored whether preferred and non-preferred pairs differed in their balance between order and complexity (different operationalizations of the balance were explored: as a sum, a difference, a multiplication, or a division). The only significant difference was when standardized order and complexity were added: on average, the sum of order and complexity for preferred images (M = 0.1770, SD = 1.2955) was higher than for non-preferred images (M = – 0.2806, SD = 1.2955), $t(59)$ = – 3.2647, p = .002. However, as the estimated difference in balance was lower than the estimated difference in order itself, this finding does not really speak in favor of a balance but rather shows the significant difference in the order score. Also, all the estimated differences of the reported t-tests were smaller than 0.52 of a standard deviation absolute difference in the standardized image measures, indicating that on average there were only very slight differences.

Magnitude of the difference between on average preferred and non-preferred images. When a preference coding was used (i.e., the image in each pair that is preferred by most participants is coded as id1), the proportion of preference for the image that was preferred by most participants correlated positively with the difference score in soothingness between the images in the pairs (r = .34, p = .0006). The bigger the standardized average soothingness rating for an image in comparison to the soothingness rating for the other image in the pair, the more participants preferred the most soothing image in the pair (see Supplementary Figure S3.a in Appendix D). The correlation between the preference proportion for the image that was preferred most often in a pair also correlated rather positively with the difference in fascination ratings between the images (r = .26, p = .0089). The bigger the standardized average fascination rating for an image in comparison to the rating for the other image in the pair, the more participants preferred the most fascinating image in the pair (see Supplementary Figure S3.b in Appendix D). No other correlations between the preference proportion for the most preferred image and the image pair measures were significant (neither for the objective statistical image properties, nor for the calculated measures of subjective balance between order and complexity). All correlations of the different image pair measures with the proportion of preference for the on average preferred image are reported in Supplementary Figure S4 in Appendix D.

Preferences in pairs with differences in order, complexity, the combination of order and complexity, soothingness, and fascination. Of the 55 pairs in which there was a difference in standardized perceived order of more than half of a standard deviation between the images in the pair, in 67.27% of the cases (i.e., 37 times) the most ordered image in the pair was preferred. In 24 of the 55 pairs the most ordered image was significantly preferred (p < .001) above the less ordered one. In 12 of the 55 pairs the least ordered image was significantly preferred (p < .001) above the most ordered one.

Of the 38 pairs in which there was a difference in standardized perceived complexity of more than half of a standard deviation between the images in the pair, in half of the cases (i.e., 19 times) the most complex image in the pair was preferred. In 11 of the 38 pairs the most complex image was significantly preferred, in 11 of the 38 pairs the least complex image was significantly preferred.

Of the 61 pairs in which there was a difference in the sum of standardized perceived order and complexity of more than half of a standard deviation between the images in the pair, in 68.85% of the cases (i.e., 42 times) the image in the pair with the highest order plus complexity score was preferred. In 26 of the 61 pairs the image with the highest combined score was significantly preferred (p < .001) above the image with the lower combined score. In 10 of the 61 pairs the image with the lowest combined score was significantly preferred (p < .001) above the image with the higher combined score.

Of the 48 pairs in which there was a difference in standardized perceived soothingness of more than half of a standard deviation between the images in the pair, in 77.08% of the cases (i.e., 37 times) the most soothing image was preferred. In 28 of the 48 pairs the most soothing image was significantly preferred above the less soothing one. In 1 of the 48 pairs the least soothing image was significantly preferred above the more ordered one. In this pair the least soothing image was slightly more fascinating than the most soothing image (0.34 of a standard deviation).

Of the 45 pairs in which there was a difference in standardized perceived fascination of more than half of a standard deviation between the images in the pair, in 84.44% of the cases (i.e., 38 times) the most fascinating image was preferred. In 29 of the 45 pairs the most fascinating image was significantly preferred above the less fascinating one. In 2 of the 45 pairs the least fascinating image was significantly preferred above the more fascinating one. In both of these pairs the least fascinating image was more than half of a standard deviation more soothing than the most fascinating image.

Of the 13 pairs in which none of these differences in order, complexity, the combination of order and complexity, soothingness, or fascination was higher than half of a standard deviation between the images in the pair, 8 pairs showed a significant difference in preference between the images in the pair.

Relations between preferences for order and complexity per individual and person properties. An individual's preference for the most ordered image in the pairs correlated positively with the individual's Personal Need for Structure (Pearson product-moment correlation coefficient $r = .29$, $p < .0001$) and with the individual's SOAQ score ($r = .29$, $p < .0001$). The higher an individual's scores on the PNS and the SOAQ, the more often the individual chose the most ordered image in the image pairs. Additionally, scoring high on Openness to Experience was related rather negatively with a preference for the most ordered image in the pairs ($r = –.16$, $p = .0012$). The correlations between a preference for order and the other Big Five personality traits (i.e., Conscientiousness, Extraversion, Agreeableness, and Neuroticism) as well as age were non-significant at the $\alpha = .01$–level.

A preference for the most complex image in the pairs correlated negatively with an individual's PNS ($r = –.20$, $p < .0001$) and with the individual's SOAQ score ($r = –.30$, $p < .0001$). The higher an individual's scores on the PNS and the SOAQ, the less often the individual chose the most complex image in the pairs. Additionally, a preference for the most complex image in the pairs was related positively with Openness to Experience ($r = .21$, $p < .0001$) and negatively with Conscientiousness ($r = –.13$, $p = .0073$). The correlations between a preference for complexity and the other Big Five personality traits (i.e., Extraversion, Agreeableness, and Neuroticism) were non-significant at the $\alpha = .01$–level. Age correlated negatively with a preference for complexity ($r = –.21$, $p < .0001$).

An overview of all correlations can be found in Figure 5. In this correlation table, also the relations between the individual difference variables are shown. As expected, individual scores on the PNS correlated positively with the individual SOAQ scores ($r = .44$, $p < .0001$) and rather negatively with Openness to Experience ($r = –.31$, $p < .0001$). However, the correlation between SOAQ scores and Openness to Experience was only marginally significant ($r = –.12$, $p = .0143$). Conscientiousness showed a slight positive correlation with both PNS ($r = .20$, $p < .0001$) and SOAQ scores ($r = .23$, $p <$

.0001). Extraversion (r = –.17, p = .0005) and Neuroticism (r = .25, $p < .0001$) only showed correlations with PNS scores, not with SOAQ scores (r = .09, p = .0823 and r = .03, p = .5372 for Extraversion and Neuroticism respectively).

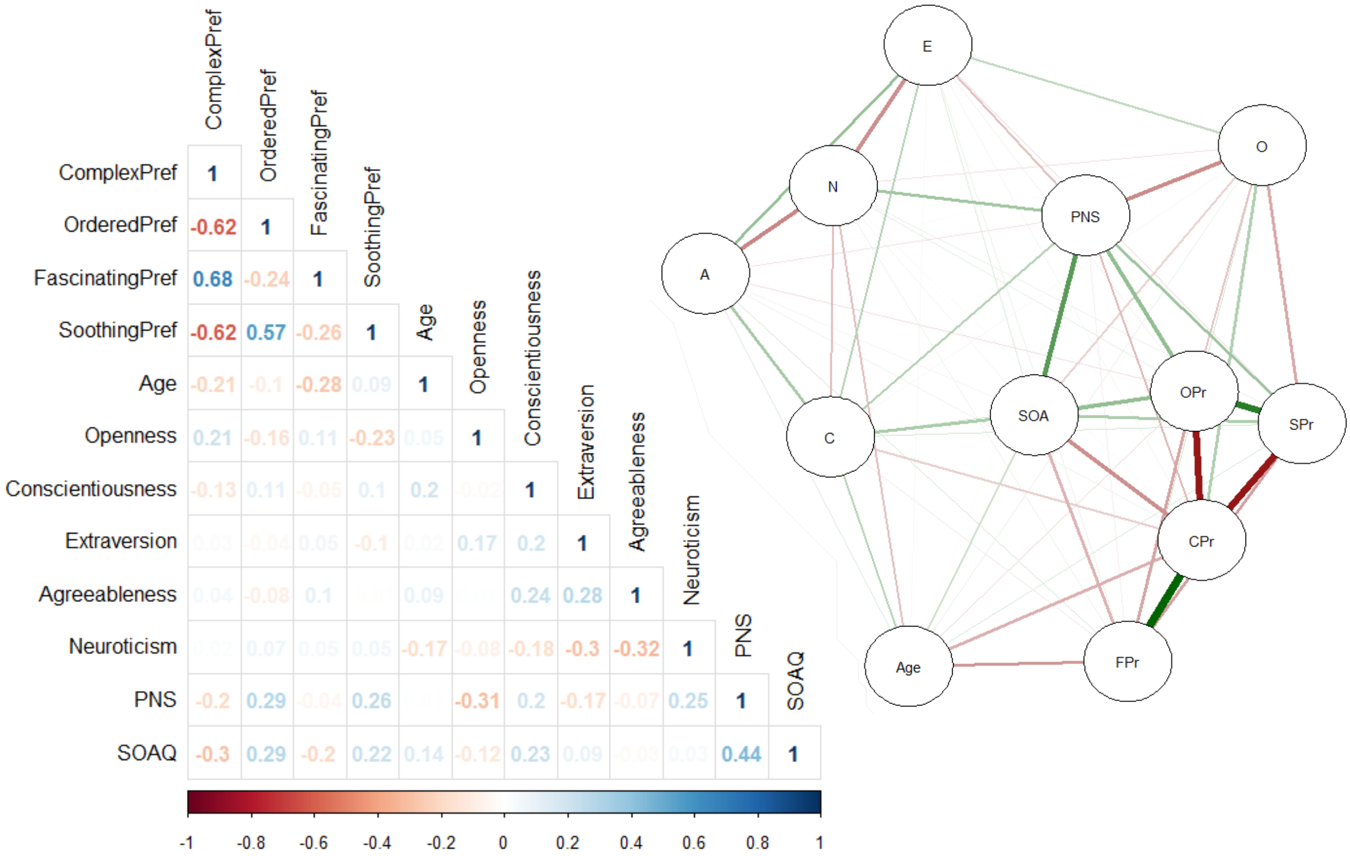


Figure 5. Correlations between the different participant measures (N = 415). Correlation plot generated with the R package corrplot (Wei & Simko, 2016), network visualization of the correlations generated with the R package qgraph (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012).

Note. OPr = preference for ordered images, CPr = preference for complex images, SPr = preference for soothing images, FPr = preference for fascinating images, O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism, PNS = Personal Need for Structure, SOAQ = Symmetry, Ordering, and Arranging Questionnaire.

Predicting individual preferences for order and complexity based on person properties: personality and age. We conducted several single and multiple linear regression analyses to explore the predictive value of several person properties and age for the individual proportions of preference for the most ordered image or the most complex image in the pairs.

For predicting individual preference for the most ordered image, the final model included the mean PNS score as a significant predictor (β = .02, $p < .0001$). This predictor explained 7.98% of the variance in preference for order, *Adj.* R^2 = .0798, $F(1,413)$ = 36.92, $p < .0001$. The higher the individual scored on the PNS, the more often this individual preferred the most ordered image in the pairs.

For predicting individual preference for the most complex image, the final model included the SOAQ sum score (β = –.024, p < .0001, *partial* R^2 = .0656), Openness to Experience (β = .019, p < .0001, *partial* R^2 = .0405), and age (β = –.018, p < .0001, *partial* R^2 = .0388) as significant predictors. Together, these predictors explained 14.74% of the variance in preference for complexity, *Adj.* R^2 = .1474, $F(3,411)$ = 24.86, p < .0001. The higher the individual scored on Openness to Experience, the more often this individual preferred the most complex image in the pairs. The higher the individual's SOAQ score and the older the individual, the less often this individual preferred the most complex image in the pairs.

Relations between preferences for soothingness and fascination per individual and person properties. A preference for the most soothing image in the pairs correlated positively with an individual's PNS (r = .26, p < .0001) and with the individual's SOAQ score (r = .22, p < .0001). The higher an individual's scores on the PNS and the SOAQ, the more often the individual chose the most soothing image in the pairs. Additionally, a preference for the most soothing image in the pairs was related negatively with Openness to Experience (r = –.23, p < .0001). The correlations between a preference for soothing images and the other Big Five personality traits (i.e., Conscientiousness, Extraversion, Agreeableness, and Neuroticism) as well as age were non-significant at the α = .01–level.

A preference for the most fascinating image in the pairs correlated negatively with the individual's SOAQ score (r = –.20, p < .0001). The higher an individual's scores on the SOAQ, the less often the individual chose the most fascinating image in the pairs. Additionally, a preference for the most fascinating image in the pairs was related negatively with age (r = –.28, p < .0001). The correlations of a preference for fascinating images with the PNS score and with the Big Five personality traits (i.e., Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) were non-significant at the α = .01–level.

Predicting individual preferences for soothingness and fascination based on person properties: personality and age. We conducted several single and multiple linear regression analyses to explore the predictive value of several person properties and age for the individual proportions of preference for the most soothing image or the most fascinating image in the pairs.

For predicting individual preference for the most soothing image, the final model included Openness to Experience (β = –.011, p = .0005, *partial* R^2 = .0289), SOAQ score (β = .009, p = .008, *partial* R^2 = .0168), and mean PNS score (β = .009, p = .009, *partial* R^2 = .0166) as significant predictors. Together, these predictors explained 10.04% of the variance in preference for order, *Adj.* R^2 = .1004, $F(3,411)$ = 16.41, p < .0001. The higher the individual scored on the PNS or the SOAQ and the lower on Openness, the more often this individual preferred the most soothing image in the pairs.

For predicting individual preference for the most fascinating image, the final model included the SOAQ sum score (β = –.013, p = .0004, *partial* R^2 = .0297) and age (β = –.020, p < .0001, *partial* R^2 = .0673) as significant predictors. Together, these predictors explained 10.13% of the variance in preference for fascination, *Adj.* R^2 = .1013, $F(2,412)$ = 24.34, p < .0001. The higher the individual's SOAQ score and the older the individual, the less often this individual preferred the most fascinating image in the pairs.

Order and Complexity Ratings for Neatly Organized Compositions

Figure 6. The 25% least complex (top) and the 25% most complex (bottom) images that are also in the 25% least ordered (left) versus the 25% most ordered (right) images of the dataset.

Amount of variability between images. There was considerable variability in the mean ratings for order between images (see Figure 6): The mean order ratings per image varied between 2.36 and 5.99 on a scale from 1 (*not at all ordered*) to 7 (*extremely ordered*; M = 4.55, SD = 0.72). The mean complexity ratings per image varied between 1.79 and 5.71 (M = 3.50, SD = 0.90). Additionally, individuals differed in the order and complexity ratings they gave for each image, and the amount of individual variation also differed between images. The standard deviations for the order ratings per image varied between 0.83 and 1.98 (M = 1.45, SD = 0.21), those for the complexity ratings between 1.10 and 1.85 (M = 1.53, SD = 0.14). Based on these results, we can conclude that there is variation between images in how ordered or complex they are perceived to be and in how much individual variation there is for each image's order or complexity score.

Amount of variability between individuals. There was considerable variability in the mean ratings for order between individuals: The mean order ratings per individual varied between 2.26 and 6.74 on a scale from 1 (*not at all ordered*) to 7 (*extremely ordered*; $M = 4.55$, $SD = 0.77$). The mean complexity ratings per individual varied between 1.46 and 6.97 ($M = 3.50$, $SD = 0.91$). Additionally, individuals differed in the order and complexity ratings they gave for each image, and the amount of variation across images also differed between individuals. The standard deviations for the order ratings per individual varied between 0.43 and 2.48 ($M = 1.38$, $SD = 0.42$), those for the complexity ratings between 0.27 and 2.24 ($M = 1.48$, $SD = 0.37$). Based on these results, we can conclude that there is variation between individuals in how ordered and complex they rated the images in general and in how much their order and complexity ratings varied across images.

Order and complexity on different perceptual stimulus dimensions. A coding scheme indicating which image pairs were selected to differ on which dimensions is included as Appendix B. Although we could not test whether the images in the pairs indeed differed in order and/or complexity on the stimulus dimension on which they were assumed to differ, we could check in how many pairs the image that was assumed to be less ordered/more complex (i.e., id1) was indeed perceived as less ordered/more complex in general than the other image in the pair (i.e., id0). In 76 of the 100 image pairs, id1 was indeed perceived as the most complex image in the pair. In 65 of the 100 image pairs, id1 was perceived as less ordered than id0.

Relations between order and complexity ratings and other image measures. The correlations between the image measures are reported in Figure 7 and in Supplementary Figure S5 in Appendix D. To calculate these correlations, standardized versions of the image measures were used (cf. "Data Analysis").

Average order rating per image. The average standardized order rating per image (averaged across participants of TON2) did not correlate significantly with the average complexity rating per image ($r = –.07$, $p = .3395$). However, the average standardized order rating did correlate positively with Fourier slope ($r = .26$, $p = .0003$) and negatively with anisotropy ($r = –.24$, $p = .0011$). No other statistical image properties correlated significantly with the standardized order ratings per image at the $\alpha = .01$–level. Furthermore, the average order rating per image was highly correlated with the average ratings of how soothing ($r = .59$, $p < .0001$) and how fascinating ($r = .54$, $p < .0001$) the image was perceived to be.

Average complexity rating per image. The average complexity rating per image (averaged across participants of TON2) did correlate significantly (*p*s < .01) with all statistical image properties in the expected directions (see Figure 7). Average complexity ratings correlated especially high with HOG complexity ($r = .62$, $p < .0001$) and fractal dimension ($r = .63$, $p < .0001$). Scatter plots for these correlations are shown in Supplementary Figure S6 in Appendix D. Furthermore, the average complexity rating per image was correlated negatively with the average ratings of how soothing ($r = –.25$, $p = .0008$) and positively with how fascinating ($r = .52$, $p < .0001$) the image was perceived to be.

Figure 7. Correlations between average rating scores per image and other image measures (N = 184; all standardized scores).

Relations between order and complexity ratings per individual and person properties.

Average order and complexity rating per individual. Average individual order ratings correlated positively with average complexity (r = .29, p = .0074), average soothingness (r = .39, p = .0002), and average fascination ratings (r = .34, p = .0016). Average individual complexity ratings also correlated positively with average soothingness (r = .28, p = .0102), and average fascination ratings (r = .52, p < .0001). In other words, individuals who gave high average ratings on one scale, also gave high average ratings on the other scales. There were no significant correlations between the order and complexity ratings per individual and other participant measures at the α = .01–level. Therefore, we decided to not conduct multiple regression analyses to predict order and complexity ratings based on person properties. All correlations are reported in Supplementary Figure S7 in Appendix D.

Soothingness and Fascination Ratings for Neatly Organized Compositions.

Amount of variability between images. There was considerable variability in the mean ratings for soothingness between images: the mean soothingness ratings per image varied between

2.10 and 5.35 on a scale from 1 (*not at all soothing*) to 7 (*extremely soothing*; $M = 3.36$, $SD = 0.74$). The mean fascination ratings per image varied between 2.26 and 5.08 ($M = 3.50$, $SD = 0.63$). Additionally, individuals differed in the soothingness and fascination ratings they gave for each image, and the amount of individual variation also differed between images. The standard deviations for the soothingness ratings per image varied between 1.13 and 1.81 ($M = 1.55$, $SD = 0.13$), those for the fascination ratings between 1.25 and 1.97 ($M = 1.63$, $SD = 0.13$). Based on these results, we can conclude that there is variation between images in how soothing or fascinating they are perceived on average, but also in how much individual variation there is for each image in how soothing or fascinating it is perceived.

Amount of variability between individuals. There was considerable variability in the mean ratings for soothingness between individuals: the mean soothingness ratings per individual varied between 1.04 and 5.29 on a scale from 1 (*not at all soothing*) to 7 (*extremely soothing*; $M = 3.36$, $SD = 0.83$). The mean fascination ratings per individual varied between 1.04 and 5.36 ($M = 3.55$, $SD = 0.87$). Additionally, individuals differed in the soothingness and fascination ratings they gave for each image, and the amount of variation across images also differed between individuals. The standard deviations for the soothingness ratings per individual varied between 0.26 and 2.39 ($M = 1.45$, $SD = 0.41$), those for the fascination ratings between 0.27 and 2.45 ($M = 1.47$, $SD = 0.39$). Based on these results, we can conclude that there is variation between individuals in how highly soothing or fascinating they indicated to perceive the images in general and in how much their soothingness and fascination scores differed between images.

Relations between soothingness and fascination ratings and other image measures.

Average soothingness ratings per image. The average rating of how soothing an image was perceived to be correlated positively with the average fascination rating per image ($r = .61$, $p < .0001$; see Figure 7). Additionally, the average soothingness rating correlated negatively with anisotropy ($r = –.29$, $p < .0001$) and rather negatively with HOG complexity ($r = –.19$, $p = .0090$). As reported before, the soothingness rating per image correlated positively with the order rating per image ($r = .59$, $p < .0001$) and negatively with the complexity rating per image ($r = –.25$, $p = .0008$). Scatter plots for those relations are shown in Figure 8. Correlations between soothingness ratings and other image measures were non-significant at the $\alpha = .01$–level.

Average fascination ratings per image. The average rating of how fascinating an image was perceived to be correlated positively with both order ($r = .54$, $p < .0001$) and complexity ratings ($r = .52$, $p < .0001$), as reported before (see Figure 7). Additionally, the average fascination rating showed similar correlations to the objective image measures as the average complexity rating: positive correlations with self-similarity ($r = .22$, $p = .0029$), HOG complexity ($r = .34$, $p < .0001$), Fourier slope ($r = .44$, $p < .0001$), and fractal dimension ($r = .47$, $p < .0001$), and a negative correlation with anisotropy ($r = –.47$, $p < .0001$). However, the correlation between anisotropy and fascination was more strongly negative than the correlation between anisotropy and average perceived complexity (significant difference at the α = .01–level). Scatter plots for those relations are shown in Figure 9.

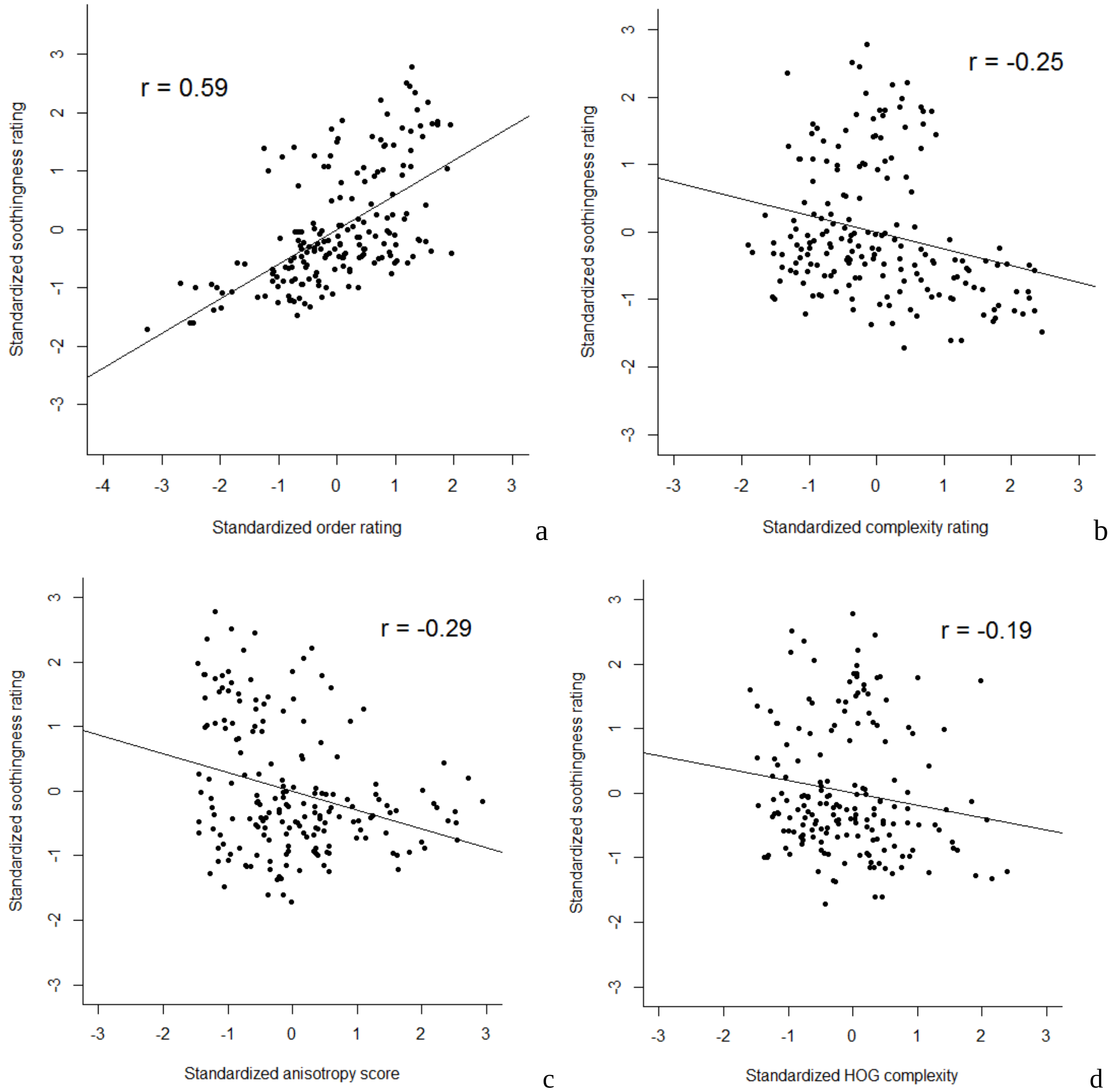


Figure 8. Scatter plots of the soothingness ratings for each image plotted against the order (a) and complexity ratings (b) for each image, as well as against anisotropy (c) and HOG complexity (d; all standardized scores).

Predicting average soothingness and fascination ratings based on measures of order and complexity.

Predicting soothingness ratings per image. We conducted multiple linear regression analyses to examine the joined impact of standardized average order and complexity ratings on the standardized average soothingness and fascination ratings. The standardized average rating of how soothing participants perceived an image to be could be significantly predicted by how ordered and how complex the image was perceived to be: the two predictors explained 38.62% of the variance, *Adj.* R^2

= .3862, $F(2,181) = 58.58$, $p < .0001$. Both the standardized average order rating and the standardized average complexity rating of the image significantly predicted its soothingness. The higher the standardized average order rating, the higher the standardized soothingness rating for that image (β = .55, $p < .0001$, *partial* $R^2 = .3539$). The lower the standardized average complexity rating, the higher the standardized soothingness rating for that image ($\beta = -.17$, $p = .0005$, *partial* $R^2 = .0643$). The model including an interaction between order and complexity did show similar significant main effects, but no significant interaction between the standardized order and complexity ratings.

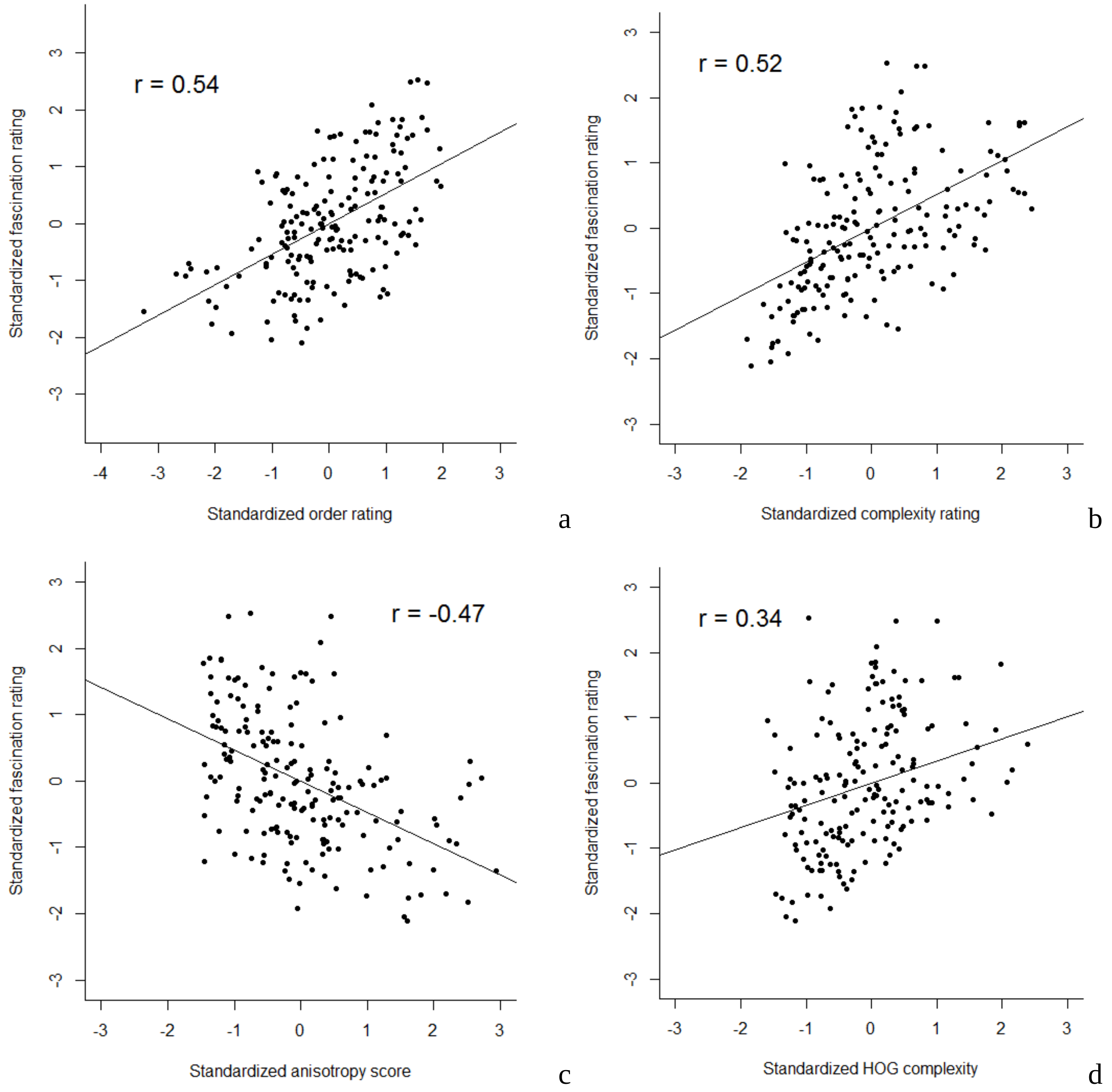


Figure 9. Scatter plots of the fascination ratings for each image plotted against the order (a) and complexity ratings (b) for each image, as well as against anisotropy (c) and HOG complexity (d; all standardized scores).

Predicting fascination ratings per image. We conducted a second multiple linear regression to test if the standardized average order and complexity ratings significantly predicted the standardized average fascination ratings per image. The two predictors explained 59.66% of the variance in fascination ratings, *Adj.* $R^2 = .5966$, $F(2,181) = 136.30$, $p < .0001$. Both the standardized average order rating ($\beta = .45$, $p < .0001$, *partial* $R^2 = .4535$) and the average complexity rating ($\beta = .39$, $p < .0001$, *partial* $R^2 = .4392$) of the image significantly predicted the fascination ratings. The higher the standardized average order rating and the higher the standardized average complexity rating for the image, the higher the standardized fascination rating was for that image (see Figure 10). The model including an interaction between order and complexity did show similar significant main effects, but the interaction was only marginally significant ($\beta = .13$, $p = .06$).

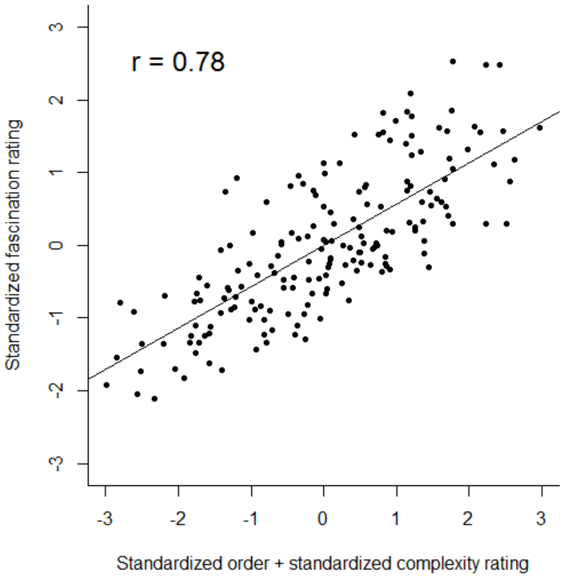


Figure 10. Scatter plot of the fascination ratings for each image plotted against the combined order and complexity ratings for each image (standardized scores, but combination not standardized).

Predicting soothingness and fascination ratings based on the balance between order and complexity. The regression models above indicated main effects of order and complexity on the average soothingness and fascination ratings given for each image. Although the interaction effect between order and complexity was not significant in the case of soothingness and only marginally significant in the case of fascination, we wanted to explore the possible influence of the balance between order and complexity on the perception of soothingness and fascination of the images in other ways.

Differential correlation between order and soothingness/fascination depending on level of complexity? For images with a positive standardized score on perceived complexity ($r = .65$, $p < .0001$), the soothingness ratings tended to increase more strongly with increasing order than for images with a negative standardized score on perceived complexity ($r = .53$, $p < .0001$), but the difference in correlation for the two pair types was not significant at the $\alpha = .01$-level (see also Figure 11a and 11b).

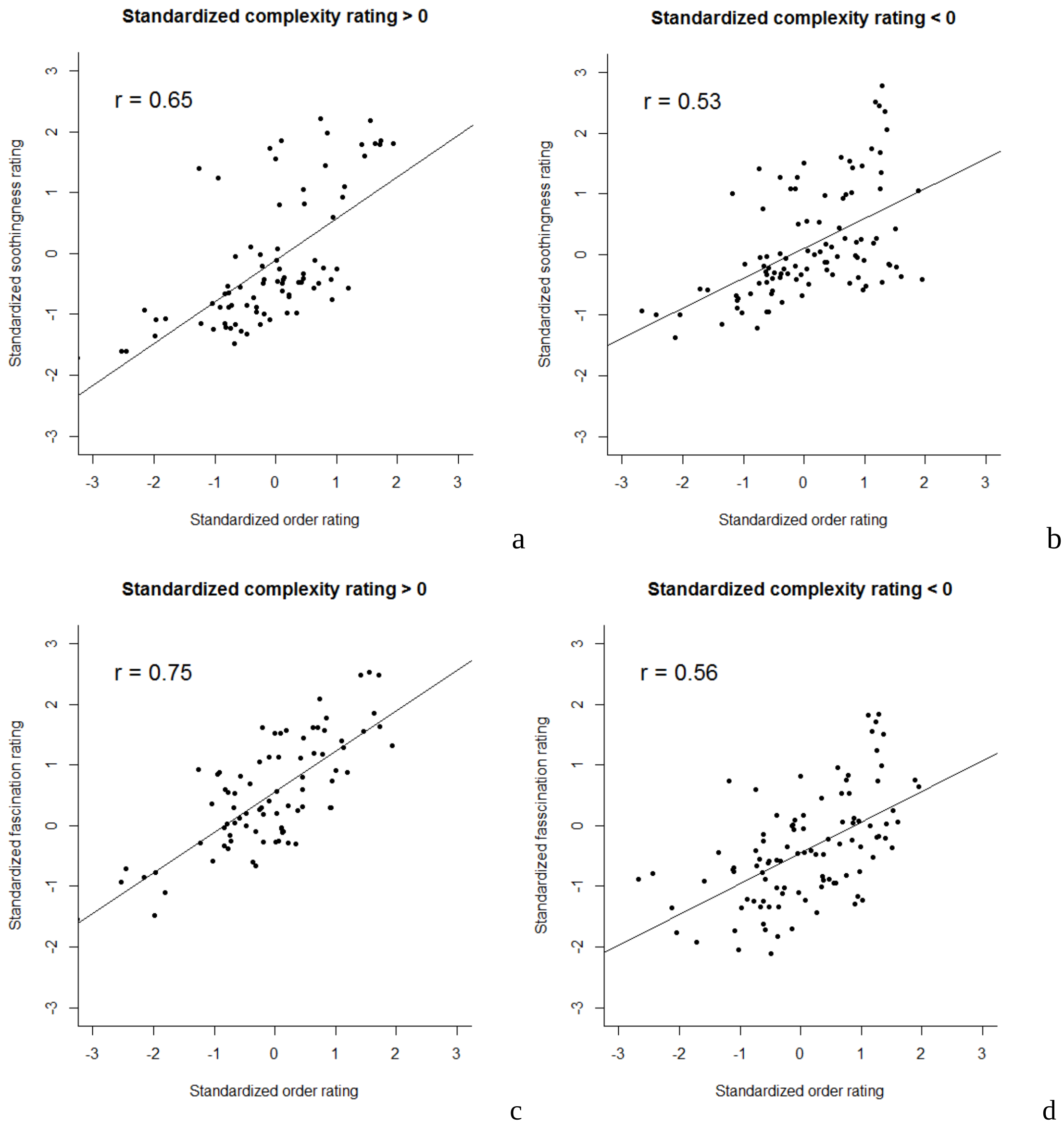


Figure 11. Scatter plots of the standardized soothingness (top) and fascination (bottom) ratings for each image plotted against the standardized order and complexity ratings for each image, separately for images with a positive (left) and a negative (right) standardized complexity rating.

For images with a positive standardized score on perceived complexity (r = .75, p < .0001), the fascination ratings increased more strongly with increasing order than for images with a negative standardized score on perceived complexity (r = .56, p < .0001). This difference in correlation between the two pair types was significant at the α = .01-level (see Figure 11c and 11d).

Relations between soothingness and fascination ratings per individual and person properties. Except for the correlation between the average soothingness rating per individual and the individual's SOAQ score (r = .28, p = .0097), no other correlations between the average ratings and other participant measures were significant at the α = .01–level. Therefore, we decided to not conduct linear regression analyses to predict soothingness and fascination ratings based on person properties.

Future Analyses to be Conducted

In the future, it would be interesting to look at individual differences in the relation between the different concepts measured in this study. To further explore relations between aesthetic preferences and interactions between person and image (pair) properties, hierarchical binary logistic regression on the actual preferences per image per person (not averaged across participants or across image pairs) can be used. For all participants to the first part of the study, the average ratings per image and the statistical image properties can be used as predictors. For all participants to the second part of the study, also the individual ratings of order, complexity, soothingness, and fascination can be used.

Although the measured personality aspects cannot predict overall differences in ratings of order, complexity, soothingness, and fascination, it could be that person properties interact with stimulus aspects in determining these ratings. To explore this, hierarchical ordered linear regression on the actual order, complexity, soothingness, and fascination ratings (not averaged across participants or across image pairs) can be used.

Furthermore, it would be interesting to explore differential relationships between perceived order and complexity and measures of aesthetic appreciation (preferences, fascination or soothingness), possibly based on personality. For example, it could be that some person properties can predict a positive relation between complexity or order and fascination, or at least the strength of that relationship. This can be investigated by clustering participants based on these correlations (e.g., by using k-means clustering) and looking for possible personality variables as predictors.

Discussion and Conclusion

Summary of the Main Findings

Neatly organized images are aesthetically pleasant. In general, images of neatly organized compositions were perceived as pleasant to look at. Most of our participants did not have extensive prior experiences with this type of images, indicating that the sample was not biased towards a subpopulation of fans of neatly organized images. We can conclude that overall, neatly organized compositions are experienced as aesthetically pleasing, even by people who do not look for this type of images regularly in daily life. This is in line with our assumption that the popularity of neatly organized images is based partly on the positive aesthetic appreciation of this type of images amongst the general population.

Both stimulus and person contribute to variability in aesthetic preferences. The amount of variability in aesthetic preferences, both between image pairs and between participants, indicated that it is worthwhile to look at both stimulus and person properties in predicting aesthetic preference.

Aesthetic preferences relate to soothingness and fascination. Concerning stimulus properties, on average preferred images were more soothing and fascinating than on average non-preferred images. In image pairs in which there was a clear difference in soothingness or fascination between the images in the pair, in almost all cases the most soothing or most fascinating image was preferred by most participants. Contrary to what was expected, balance scores between order and complexity did not clearly relate with aesthetic preferences. There was some relation between subjective order ratings and aesthetic preferences for neatly organized compositions. Which more objective stimulus properties relate to general preferences for specific neatly organized compositions should be investigated further.

High symmetry, ordering, and arranging tendencies, high personal need for structure, and low openness relate to a preference for ordered images. Concerning person properties, we found indications for individual differences in aesthetic preferences for order based on personality (e.g., Openness to Experience, symmetry, ordering, and arranging tendencies). The higher a person scored on questionnaires relating to personal need for structure and symmetry, ordering, and arranging tendencies, the more often that person chose the most ordered image in the pairs. The higher a person scored on Openness to Experience, the less often he or she preferred the most ordered image in the pairs. For predicting individual proportions of preference for the most ordered image, the final model included an individual's personal need for structure as the only predictor, explaining almost 8% of the variability in the preference proportions.

Low symmetry, ordering, and arranging tendencies, low personal need for structure, young age, and high openness relate to a preference for complex images. The lower a person's symmetry, ordering, and arranging tendencies (as indicated by the score on the SOAQ), the lower his or her personal need for structure and the higher his or her Openness to Experience, the more often that person preferred the most complex image in the pairs. Age showed a negative relation to a preference for complexity: The older the person, the less often he or she chose the most complex image in the pairs. For predicting individual proportions of preference for the most complex image in the pairs, the final model included an individual's SOAQ score, Openness to Experience, and age, together explaining almost 15% of the variability in the preference proportions.

High symmetry, ordering, and arranging tendencies, high personal need for structure, and low openness relate to a preference for soothing images. Often choosing the most soothing image in the pairs did relate positively with an individual's symmetry, ordering, and arranging tendencies and his or her personal need for structure. High openness related to less often choosing the most soothing image in the pairs. Proportions of preference for the most soothing image in the pairs thus correlated with the same person properties as did proportions of preference for the most ordered image in the pairs, although the correlations with preferences for soothingness were somewhat less strong. Possibly the correlations found with preferences for soothingness are a consequence of the correlations of the person properties with preferences for order, especially since the correlation between the two types of preferences was so high (r = .57, p < .0001).

Low symmetry, ordering, and arranging tendencies and young age relate to a preference for fascinating images. Often choosing the most fascinating image in the pairs did relate negatively with an individual's symmetry, ordering, and arranging tendencies and age. The pattern of correlations between proportions of preference for fascination and the person properties resembled that of proportions of preference for complexity. Possibly the correlations found with preferences for fascination are a consequence of the correlations of the person properties with preferences for complexity. These two types of preference correlated highly positive (r = .68, p < .0001).

Both stimulus and person contribute to variability in order and complexity ratings. The amount of variability in order and complexity ratings, both between images and between participants, indicated that it is worthwhile to look at both stimulus and person properties relating to perceptions of order and complexity.

Objective measures of complexity relate to subjective complexity. Concerning the stimulus properties, perceived complexity related highly positive with some objective indicators of complexity (i.e., HOG-based complexity, fractal dimension, and to a certain extent Fourier slope). This finding strengthens the usefulness of this type of statistical image properties as indicators of subjective complexity in the search for factors relating to aesthetic appreciation.

Unclear which person properties relate to overall perceptions of order and complexity. Although considerable variability existed in the average order and complexity ratings per individual, the person properties measured in this study did not associate with individual differences in mean order and complexity ratings. Differences in the overall perceived level of order and complexity could thus not explain the relation between person properties and overall aesthetic preferences for order and complexity. Probably, interactions between stimulus and person in the perceptions of order and complexity are more interesting to explore than the main effect of a person’s characteristics across all images.

Both stimulus and person contribute to variability in soothingness and fascination ratings. As for order and complexity, considerable variation was found in the average soothingness and fascination ratings, both per image and per individual. It is thus worthwhile to explore both stimulus and person properties relating to perceptions of soothingness and fascination.

Soothingness relates positively to subjective order and negatively to subjective complexity. Soothingness ratings related positively with order but slightly negative with complexity. In predicting the average soothingness score for an image, order was by far the most important predictor, but complexity still significantly increased the model fit. Together, order and complexity ratings explained almost 39% of the variability in the average soothingness ratings per image.

Fascination relates positively to both subjective order and subjective complexity. Fascination showed independent positive associations with both order and complexity. In predicting the average fascination ratings per image, both order and complexity were significant predictors, together explaining almost 60% of the variability in the average fascination ratings per image.

Unclear which person properties relate to overall perceptions of soothingness and fascination. Although considerable variability existed in the average soothingness and fascination ratings per individual, the person properties measured in this study did not strongly associate with individual differences in mean soothingness and fascination ratings. Only symmetry, ordering, and arranging tendencies showed a slight correlation with mean soothingness ratings per individual. The higher a person's tendencies for symmetry, ordering, and arranging, the higher his or her mean soothingness rating across all images. Probably, interactions between stimulus and person in the perceptions of soothingness and fascination are more interesting to explore than the main effect of a person's characteristics across all images.

Theoretical Reflections

Based on the literature review and the results of this study, we propose different possible pathways towards aesthetic appreciation, influenced by both stimulus and person characteristics, and especially the interaction between them (see Figure 12).

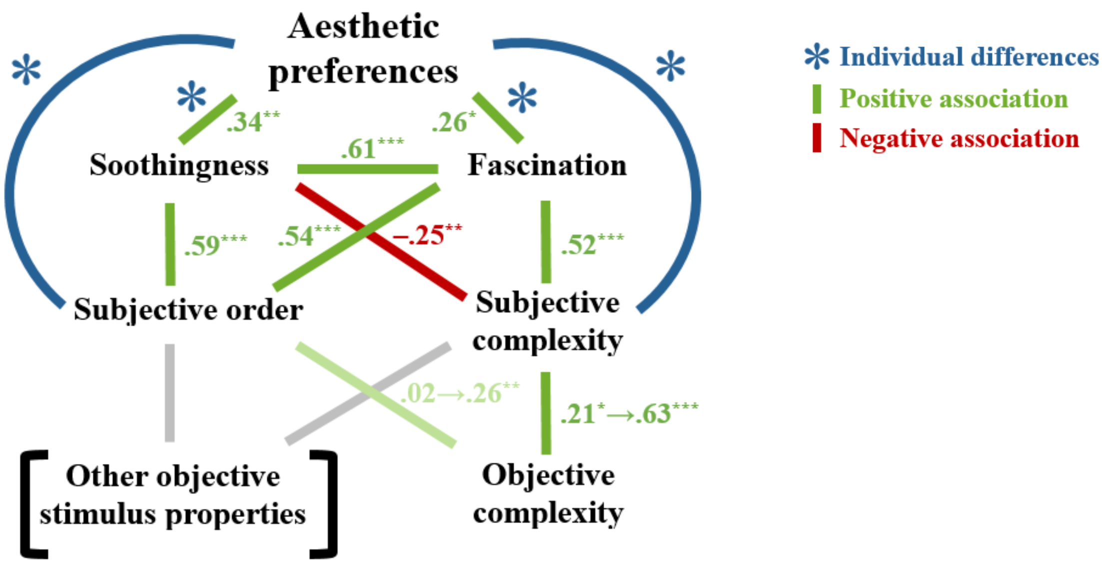


Figure 12. Schematic representation of the study's main findings.
Note. Numbers indicate Pearson product-moment correlations; *p < .01, **p < .001, ***p < .0001.

When investigating images of neatly organized compositions, aesthetic preferences relate positively with both the perceived soothingness and fascination of the images. However, soothingness and fascination relate differently to the perceived order and complexity of the image. Whereas the soothingness of an image can be predicted by high order and low complexity, how fascinating an image is perceived to be is associated with high order and high complexity. Individuals differ in the extent to which their aesthetic preferences are associated with perceived order and complexity, and (consequently) also in the extent to which their aesthetic preferences are associated with soothingness and fascination[14]. Symmetry, ordering, and arranging tendencies, a personal need for structure, and

[14] Preferences for soothingness correlated with the same person properties as did preferences for order, although somewhat less strongly so. Also the overall pattern of correlations of the person properties was similar for preferences for fascination and preferences for complexity. Possibly the correlations found with preferences for soothingness and fascination are a consequence of the correlations of the person properties with preferences for order and complexity. The observation that also the correlations of soothingness and order ratings per image with

low openness are indicators of a preference for the most ordered images in the pairs, whereas low symmetry, ordering, and arranging tendencies, low personal need for structure, high openness, and young age related to a preference for the most complex images in the pairs.

Some objective indicators of complexity (i.e., HOG-based complexity, fractal dimension, and to a certain extent Fourier slope) related strongly positively with subjective complexity. Some objective indicators of complexity showed a slightly positive relation with subjective order, but these relations should be interpreted rather cautiously.

In future analyses and research, it would be interesting to further investigate (a) which person characteristics relate to individual differences in the relations of soothingness and fascination with order and complexity; and (b) which other objective stimulus characteristics relate to subjective order and complexity.

Different routes to aesthetic appreciation. Although soothingness and fascination differ in their relations with order and complexity, both soothingness and fascination seemed to relate positively with aesthetic preferences. These findings can be related to the Pleasure-Interest Model of aesthetic appreciation proposed by Graf and Landwehr (2015, 2017), in which aesthetic appreciation can be mediated by both pleasure and interest. We assume pleasure-based liking to relate to preferences for soothingness and interest-based liking to preferences for fascination. In future analyses and studies, it would be interesting to investigate individual differences in the relative importance of soothingness and fascination in predicting aesthetic preferences, and relatedly also differences in processing styles (i.e., automatic or controlled), as suggested in the model of Graf and Landwehr (2015, 2017).

Combination of rather than balance between order and complexity. Like in the literature, the relation between order and complexity seemed to be both complementary and antagonistic. On the one hand, order and complexity *complement* each other in predicting how fascinating an image is perceived to be. On the other hand, order and complexity are *partial opposites* in their relation to soothingness. The balance between order and complexity seems to be a *combination* of the independent relations of order and complexity with different types of aesthetic appreciation, rather than an *interaction* between order and complexity. Our findings for fascination also relate to the findings of Post et al. (2016), who stated that it is the maximization of both unity (i.e., order) and variety (i.e., complexity) that yields the greatest aesthetic appreciation. Future analyses and studies could focus on individual differences in the relation between order and complexity and the relation of order and complexity with different types of aesthetic appreciation.

---

the statistical image properties were similar, as well as those of fascination and complexity ratings per image with the statistical image properties, strengthens that idea. In the same vein, Lyssenko et al. (2016) found interest ratings to show similar associations with self-similarity, HOG complexity, and anisotropy as complexity ratings, but the associations were less strong for interest than for complexity.

Limitations

Specific type of stimuli. The observation that different types of positive aesthetic appreciation correlated positively with each other can possibly be explained based on their common positive association with order. Overall, order had a stronger relation with the different types of aesthetic appreciation assessed (i.e., preferences, soothingness, fascination). This finding is probably heavily influenced by the selected stimulus type, and should therefore be interpreted with the necessary caution.

Low experimental control. By selecting existing images of neatly organized compositions for this study but only matching images that were assumed to be comparable, we attempted to balance the needs for ecological validity and experimental control. However, the different image measures correlated quite highly, which makes it difficult to independently investigate the relations of the different measures with aesthetic appreciation. In future studies, it would be good to have both more ecologically valid and more controlled investigations, to increase relevance of the findings for daily life on the one hand and to increase knowledge about the causal relationships between the measured concepts on the other hand.

Exploratory nature of the project. Although we specified research questions and hypotheses at the start of the research (cf. "Research questions and hypotheses"), not all of them were very clearly specified (e.g., the type of balance between order and complexity) and the main aim of this study was to explore possible interesting relationships between the measured concepts. There is need for additional investigations on newly collected data to confirm the relationships found on the basis of this research.

Practical aspects of the study. As the data were collected online, we had no control over the specific context in which participants completed the study. In the first part of the study, the image in each pair that was shown left or right coincided with the image that was assumed to be least or most complex. Although this was counterbalanced between participants, for each individual participant this left or right bias cannot be disentangled from the bias for the image that was assumed to be most complex. However, across all image pairs, image position did not significantly affect preference proportions for that image. Additionally, more objective codings of the image pairs were used for most of the analyses, which did not coincide with the image position in the pair.

Participation times as well as oral communication with some participants indicated that the second part of the study was long and tiring. Although it might have been better for participants if the second part of the study did not ask ratings for all images used in the first part, it would also decrease the amount of collected information. Another possibility would have been to allow participants to pause the study and finish it on a later moment in time.

Many analysis options. Although we predicted an influence of the mentioned person properties on both absolute preferences for order and/or complexity and on relative levels of order in proportion to the level of complexity, we did not (yet) test the influence of the person properties on the relative levels, as the other analyses on the balance measures indicated that order and complexity seem to have rather independent relations with other types of aesthetic appreciation (i.e., soothingness and fascination). Also many other analysis options were not explored yet, because of the abundance of possibilities with the collected data, as it has many different possible dependent variables and units of analysis. Additionally, a lot of choices had to be made while doing the analyses and reporting (e.g., coding of the image pairs, using standardized or unstandardized scores). We tried to communicate as clearly and transparently as possible about the different choices made and analyses conducted.

Other factors predicting aesthetic appreciation. Although the focus of this study was on stimulus and person properties related to (the balance between) order and complexity, many other factors might play a role in aesthetic appreciation. For example, color aspects of the image could be important, as well as whether the image represents natural or non-natural objects. Also, the mood or goal state of the participants was not measured in this study, although this could also influence aesthetic appreciation (see also motivational orientation in Deng & Poole, 2012, and the Pleasure-Interest Model of Aesthetic Liking by Graf & Landwehr, 2015, 2017). Conceptual aspects could also be relevant in some types of neatly organized compositions. Further analyses on the collected data as well as new studies could explore these factors and many more.

We found indications for many individual differences in aesthetic appreciation (i.e., preferences, soothingness, and fascination). For some of the preferences, we found a few suggestions about possible driving factors, but there is still a lot to explain. In future studies and analyses, it would be good to further explore interactions between stimulus and person properties that could relate to aesthetic appreciation.

Conclusion

In conclusion, aesthetic appreciation for images of neatly organized compositions is closely related to both the amount of order and complexity perceived in the images. Images are perceived as more fascinating as they are more ordered and more complex. Soothing images are highly ordered but low in complexity. Individuals differ in their preferences for more ordered, complex, soothing, and fascinating images, and these differences can partly be explained by differences in personality. In general, both stimulus and person interact in determining aesthetic appreciation. Further investigations should focus on both empirically investigating and theoretically explaining these interactions.

References

Adkins, O. C., & Norman, J. F. (2016). The visual aesthetics of snowflakes. *Perception, 45*, 1304–1319. doi:10.1177/2041669516661122

Aitken, P. P. (1974). Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology*, *103*, 240–244. doi:10.1037/h0036787

Alleyne, A. (2015, October 12). This is your brain on tidiness: The psychology of 'organization porn'. *CNN.* Retrieved from http://edition.cnn.com/

Arnheim, R. (2004). *Art and visual perception: A psychology of the creative eye, fiftieth anniversary printing* (Rev. ed.). Berkeley and Los Angeles, CA: University of California Press. (Original work published 1954)

Arnheim, R. (1966). Order and complexity in landscape design. In R. Arnheim (Ed.), *Toward a psychology of art: Collected essays* (pp. 123–135). Berkeley and Los Angeles, CA: University of California Press.

Arnheim, R. (1971). *Entropy and art: An essay on disorder and order*. Berkeley and Los Angeles, CA: University of California Press.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology, 53*, 221–227. doi:10.1037/h0043921

Bardi, A., Guerra, V. M., Ramdeny, G. S. D. (2009). Openness and ambiguity intolerance: Their differential relations to well-being in the context of an academic life transition. *Personality and Individual Differences, 47*, 219–223. doi:10.1016/j.paid.2009.03.003

Barla, A., Franceschi, E., Odone, F., & Verri, A. (2002). Image kernels. *Pattern Recognition with Support Vector Machines: First International Workshop, SVM 2002. Lecture Notes in Computer Science*, *2388,* 83–96. doi: 10.1007/3-540-45665-1_7

Belke, B., Leder, H., & Carbon, C. C. (2015). When challenging art gets liked: Evidences for a dual preference formation process for fluent and non-fluent portraits. *PLoS ONE*, *10,* e0131796. doi:10.1371/journal.pone.0131796

Berlyne, D. E. (1960). *Conflict, arousal and curiosity.* New York, NY: McGraw-Hill.

Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology, 22*, 376–387. doi:10.1037/h0082777

Bielski, Z. (2010, November 1). Organization as art. *The Globe and Mail.* Retrieved from http://www.theglobeandmail.com/life/the-hot-button/organization-as-art/article1781265/

Birkhoff, G. D. (1933). *Aesthetic measure.* Cambridge, MA: Harvard University Press.

Bosch, A., Tisserman, A., & Munoz, X. (2007). Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval* (pp. 401–408). New York, NY: Association of Computing Machinery. doi:10.1145/1282280.1282340

Boselie, F., & Leeuwenberg, E. (1985). Birkhoff revisited: Beauty as a function of effect and means. *American Journal of Psychology*, *98*, 1–39. doi:10.2307/1422765

Berenbaum, H., Bredemeier, K., & Thompson, R. J. (2008). Intolerance of uncertainty: Exploring its dimensionality and associations with need for cognitive closure, psychopathology, and personality. *Journal of Anxiety Disorders*, *22*, 117–125. doi:10.1016/j.janxdis.2007.01.004

Braun, J., Amirshahi, S. A., Denzler, J., & Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, *4*, 808. doi:10.3389/fpsyg.2013.00808

Buhr, K., & Dugas, M. J. (2006). Investigating the construct validity of intolerance of uncertainty and its unique relationship with worry. *Journal of Anxiety Disorders*, *20*, 222–236. doi:10.1016/j.janxdis.2004.12.004

Cárdenas, R. A., & Harris, L. J. (2006). Symmetrical decorations enhance the attractiveness of faces and abstract designs. *Evolution and Human Behavior*, *27*, 1–18. doi:10.1016/j.evolhumbehav.2005.05.002

Cavazos, J. T., Judice-Campbell, N., & Ditzfeld, C. P. (2012). Differing emotional sensitivities in the two factors of personal need for structure. *Journal of Research in Personality*, *46*, 49–54. doi:10.1016/j.jrp.2011.12.005

Ceulemans, E., & Meers, K. (2005). HiClas (Version 0.9 beta) [Computer software].

Chamorro-Premuzic, T., Burke, C., Hsu, A., & Swami, V. (2010). Personality predictors of artistic preferences as a function of the emotional valence and perceived complexity of paintings. *Psychology of Aesthetics, Creativity, and the Arts*, *4*, 196–204. doi:10.1037/a0019211

Chamorro-Premuzic, T., Furnham, A., & Reimers, S. (2007). Personality and art. *The Psychologist*, *20*, 84–87. Retrieved from https://thepsychologist.bps.org.uk/

Chevrier, J., & Delorme, A. (1980). Aesthetic preferences: Influence of perceptual ability, age and complexity of stimulus. *Perceptual and Motor Skills*, *50*, 839–849. doi:10.2466/pms.1980.50.3.839

Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity of images: Experimental and computational estimates compared. *Perception*, *41*, 631–647. doi:10.1068/p6987

Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental Psychology: General*, *106*, 269–301. doi:10.1037/0096-3445.106.3.269

Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 365–378. doi:10.1037/0096-1523.5.2.365

Corchs, S. E., Ciocca, G., Bricolo, E., & Gasparini, F. (2016). Predicting complexity perception of real world images. *PLoS ONE*, *11*, e0157986. doi:10.1371/journal.pone.0157986

Crego, C., Samuel, D. B., & Widiger, T. A. (2015). The FFOCI and other measures and models of OCPD. *Assessment*, *22*, 135–151. doi:10.1177/1073191114539382

Cupchik, G. C. (1986). A decade after Berlyne: New directions in experimental aesthetics. *Poetics*, *15*, 345–369. doi:10.1016/0304-422X(86)90003-3

Cupchik, G. C., & Gebotys, R. J. (1988). The search for meaning in art: Interpretive styles and judgments of quality. *Visual Arts Research*, *14*(2), 38–50.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *International Conference on Computer Vision & Pattern Recognition*, *1*, 886–893. doi:10.1109/CVPR.2005.177

De Boeck, P., Rosenberg, S., & Van Mechelen, I. (1993). The hierarchical classes approach: A review. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 265–286). San Diego, CA: Academic Press.

Deng, L., & Poole, M. S. (2012). Aesthetic design of e-commerce web pages – Webpage Complexity, Order and preference. *Electronic Commerce Research and Applications*, *11*, 420–440. doi:10.1016/j.elerap.2012.06.004

Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, *90*, 152–157. doi:10.1080/00223890701845229

Dugas, M. J., Hedayati, M., Karavidas, A., Buhr, K., Francis, K., & Phillips, N. A. (2005). Intolerance of uncertainty and information processing: Evidence of biased recall and interpretations. *Cognitive Therapy and Research*, *29*, 57–70. doi:10.1007/s10608-005-1648-9

Ellison, K. (2015, April 8). Things Organized Neatly: Meet the internet’s tidiest blogger. Retrieved from http://www.goexplore.net/internet-heroes/things-organized-neatly-interview/

Epskamp, S., Cramer, A. O. J., Waldorp, L. J. Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18. doi:10.18637/jss.v048.i04

Eysenck, H. J. (1941). ‘Type’-factors in aesthetic judgements. *British Journal of Psychology*, *31*, 262–270. doi:10.1111/j.2044-8295.1941.tb00992.x

Eysenck, H. J. (1942). The experimental study of the 'good Gestalt'—A new approach. *Psychological Review*, *49*, 344–364. doi:10.1037/h0057013

Fechner, G. T. (1876). Vorschule der Aesthetik [Elementary aesthetics]. Leipzig, Germany: Breitkopf & Härtel.

Fletcher, N. H. (2012). The sound of music: Order from complexity. *Acoustics Australia*, *40*, 188–193.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*, 49–70. doi:10.1348/000712610X498958

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences*, *17*, 791–802. doi:10.1016/0191-8869(94)90048-5

Friedenberg, J., & Liby, B. (2016). Perceived beauty of random texture patterns: A preference for complexity. *Acta Psychologica*, *168*, 41–49. doi:10.1016/j.actpsy.2016.04.007

Gibbs, N. A. (1996). Nonclinical populations in research on obsessive-compulsive disorder: A critical review. *Clinical Psychology Review*, *16*, 729–773. doi:10.1016/S0272-7358(96)00043-8

Gilbert, K. E., & Kuhn, H. (1972). *A history of esthetics: Revised and enlarged* (Rev. ed.). Westport, CT: Greenwood Press. (Original work published 1953)

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, *48*, 82–98. doi:10.1037/0022-3514.48.1.82

Gombrich, E. H. (1992). *The sense of order: A study in the psychology of decorative art* (2nd ed.). London, England: Phaidon. (Original work published 1984)

Gordon, J., & Gridley, M. C. (2013). Musical preferences as a function of stimulus complexity of piano jazz. *Creativity Research Journal*, *25*, 143–146. doi:10.1080/10400419.2013.752303

Graf, L. K. M., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, *19*, 395–410. doi:10.1177/1088868315574978

Graf, L. K. M., & Landwehr, J. R. (2017). Aesthetic pleasure versus aesthetic interest: The two routes to aesthetic liking. *Frontiers in Psychology*, *8*, 15. doi:10.3389/fpsyg.2017.00015

Gray, K., Schmitt, P., Strohminger, N., & Kassam, K. S. (2014). The science of style: In fashion, colors should match only moderately. *PLoS ONE*, *9*, e102772. doi:10.1371/journal.pone.0102772

Green, P. (2013, March 27). The art of unjumbling. *The New York Times.* Retrieved from http://www.nytimes.com/

Grenier, S., Barrette, A.-M., & Ladouceur, R. (2005). Intolerance of Uncertainty and Intolerance of Ambiguity: Similarities and differences. *Personality and Individual Differences*, *39*, 593–600. doi:10.1016/j.paid.2005.02.014

Güçlütürk, Y., Jacobs, R. H. A. H., & van Lier, R. (2016). Liking versus complexity: Decomposing the inverted U-curve. *Frontiers in Human Neuroscience*, *10*, 112. doi:10.3389/fnhum.2016.00112

Hagerhall, C. M., Purcell, T., & Taylor, R. (2004). Fractal dimension of landscape silhouette outlines as a predictor of landscape preference. *Journal of Environmental Psychology*, *24*, 247–255. doi:10.1016/j.jenvp.2003.12.004

Hay, S. (2015, July 31). The history of knolling. Retrieved from https://web.archive.org/web/20150731150518/http://www.lyst.com/longlyst/art-knolling/

Het Nieuwe TeamWerken (2015). Persoonlijke Nood aan Structuur [Measurement instrument]. Retrieved from http://hetnieuweteamwerken.be/tools/vragenlijst-structuurbehoefte

Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, *7*, 335. doi:10.3389/fpsyg.2016.00335

Imamoglu, Ç. (2000). Complexity, liking and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology*, *20*, 5–16. doi:10.1006/jevp.1999.0155

Innis, E. (2016, February 29). Creative trends close-up: Flat lay in fashion, food, and social media. Retrieved from http://www.shutterstock.com/blog/

Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology*, *95*, 41–56. doi:10.1348/000712604322779451

Jacobsen, T., & Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills*, *95*, 755–766. doi:10.2466/pms.2002.95.3.755

Jacobsen, T., Schubotz, R. I., Höfel, L., & von Cramon, D. Y. (2006). Brain correlates of aesthetic judgment of beauty. *NeuroImage*, *29*, 276–285. doi:10.1016/j.neuroimage.2005.07.010

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (Vol. 3, pp. 114–158). New York, NY: Guilford Press.

Joye, Y., Steg, L., Ünal, A. B., & Pals, R. (2016). When complex is easy on the mind: Internal repetition of visual information in complex objects is a source of perceptual fluency. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 103–114. doi:10.1037/xhp0000105

Kein & Aber AG (2012). Tidying up Art (Version 1.0.0) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/kunst-aufraumen/id532996706

King, R., Villeneuve, E., Post, L., Flowers, C., & Moonshine, K. (1995). Aesthetic preference and DSM-IIIR personality disorders. *Personality and Individual Differences, 18,* 797–799. doi:10.1016/0191-8869(95)00012-U

Koffka, K. (1940). Problems in the psychology of art. In R. Bernheimer, R. Carpenter, K. Koffka, & M. C. Nahm (Eds.), *Art: A Bryn Mawr Symposium*. Bryn Mawr Notes and Monographs (Vol. IX, pp. 180–273). New York: Sentry Press.

Kreitler, S., Zigler, E., & Kreitler, H. (1974). The complexity of complexity. *Human Development*, *17*, 54–73. doi:10.1159/000271333

Kruglanski, A. W. (1990). Motivations for judging and knowing: Implications for causal attribution. In E. T. Higgins & R. M. Sorrentino (Eds.), *The handbook of motivation and cognition: Foundation of social behaviour* (Vol. 2, pp. 333–368). New York, NY: Guilford Press.

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

Lipps, T. (1923). Ästhetik. Psychologie des Schönen und der Kunst. Erster Teil: Grundlegung der Ästhetik [Aesthetics. Psychology of beauty and art. First part: Basics of aesthetics]. Leipzig, Germany: Leopold Voss.

Locher, P. J., Stappers, P. J., & Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychologica*, *99*, 141–161. doi:10.1016/S0001-6918(98)00008-0

Lyssenko, N., Redies, C., & Hayn-Leichsenring, G. U. (2016). Evaluating abstract art: Relation between term usage, subjective ratings, image properties, and personality traits. *Frontiers in Psychology*, *7*, 973. doi:10.3389/fpsyg.2016.00973

Mac Donald, A. P. (1970). Revised scale for ambiguity tolerance: Reliability and validity. *Psychological Reports*, *26*, 791–798. doi:10.2466/pr0.1970.26.3.791

Marin, M. M., & Leder, H. (2013). Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS ONE*, *8*, e72412. doi:10.1371/journal.pone.0072412

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

Mather, G. (2014). *The psychology of visual art: Eye, brain and art.* Cambridge, England: Cambridge University Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x

Moore, H. T. (1917). *Pain and pleasure.* New York, NY: Moffat, Yard and Company.

Munsinger, H., & Kessen, W. (1964). Uncertainty, structure, and preference. *Psychological Monographs: General and Applied*, *78*(9), 1–24. doi:10.1037/h0093865

Mureika, J. R., and Taylor, R. P. (2013). The abstract expressionists and les automatistes: a shared multi-fractal depth? *Signal Processing*, *93*, 573–578. doi:10.1016/j.sigpro.2012.05.002

Muth, C., & Carbon, C.-C. (2016). SeIns: Semantic instability in art. *Art & Perception*, *4*, 145–184. doi:10.1163/22134913-00002049

Nadal, M. (2007). Complexity and aesthetic preference for diverse visual stimuli. (Doctoral thesis, Universitat de les Illes Balears, Spain). Retrieved from http://ibdigital.uib.cat/greenstone/cgi-bin/library.cgi

Nadal, M., Munar, E., Marty, G., & Cela-Conde C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, *28*, 173–191. doi:10.2190/EM.28.2.d

Nasar, J. L. (1994). Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and Behavior*, *26*, 377–401. doi:10.1177/001391659402600305

Neuberg, S. L., & Newsom, J. T. (1993). Personal Need for Structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*, *65*, 113–131.

Nicki, R. M., & Moss, V. (1975). Preference for non-representational art as a function of various measures of complexity. *Canadian Journal of Psychology*, *29*, 237–249. doi:10.1037/h0082029

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology*, *14*, 77–93. doi:10.1037/h0094090

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*, 1041–1046. Retrieved from http://escholarship.org/uc/item/17s4h6w8

Orr, M. G., & Ohlsson, S. (2005). Relationship between complexity and liking as a function of expertise. *Music Perception: An Interdisciplinary Journal*, *22,* 583–611. doi:10.1525/mp.2005.22.4.583

Palmer, S. E., & Griscom, W. S. (2013). Accounting for taste: Individual differences in preference for harmony. *Psychonomic Bulletin & Review*, *20*, 453–461. doi:10.3758/s13423-012-0355-2

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, *64*, 77–107. doi:10.1146/annurev-psych-120710-100504

Phillips, F., Norman, J. F., & Beers, A. M. (2010). Fechner's aesthetics revisited. *Seeing and Perceiving*, *23*, 263–271. doi:10.1163/187847510X516412

Post, R. A. G., Blijlevens, J., & Hekkert, P. (2016). 'To preserve unity while almost allowing for chaos': Testing the aesthetic principle of unity-in-variety in product design. *Acta Psychologica*, *163*, 142–152. doi:10.1016/j.actpsy.2015.11.013

Post, R., Nguyen, T., & Hekkert, P. (2017). Unity in variety in website aesthetics: A systematic inquiry. *International Journal of Human–Computer Studies*, *103*, 48–62. doi:10.1016/j.ijhcs.2017.02.003

Radcliffe, A. (2016, October 15). This blog is such a nice place on the internet, thank you for curating it. [Web log post]. Retrieved from http://thingsorganizedneatly.tumblr.com/post/151846978729/this-blog-is-such-a-nice-place-on-the-internet/

Radomsky, A. S., & Rachman, S. (2004). Symmetry, ordering and arranging compulsive behaviour. *Behaviour Research and Therapy*, *42*, 893–913. doi:10.1016/j.brat.2003.07.001

Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, *6*(6–7), 15–51.

R Core Team (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364–382. doi:10.1207/s15327957pspr0804_3

Redies, C., Amirshahi, S. A., Koch, M., & Denzler, J. (2012). PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. *ECCV 2012 Workshops and Demonstrations, Part I. Lecture Notes in Computer Science*, *7583*, 522–531. doi:10.1007/978-3-642-33863-2_54

Redies, C., Brachmann, A., & Hayn-Leichsenring, G. U. (2014). Changes of statistical properties during the creation of graphic artworks. *Art & Perception*, *3*, 93–116. doi:10.1163/22134913-00002017

Redies, C., Hänisch, J., Blickhan, M., & Denzler, J. (2007). Artists portray human faces with the Fourier statistics of complex natural scenes. *Network*, *18*, 235–248. doi:10.1080/09548980701574496

Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: similarity to natural scenes. *Spatial Vision*, *21*, 137–148. doi:10.1163/156856807782753921

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*, 1236–1256. doi:10.1037/0022-3514.84.6.1236

Rist, J. M. (1967). Plotinus: *The road to reality*. Cambridge, England: Cambridge University Press.

Rosenberg, S., Van Mechelen, I., & De Boeck, P. (1996). A hierarchical classes model: Theory and method with applications in psychology and psychopathology. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), Clustering and classification (pp. 123–155). River Edge, NJ: World Scientific Publishing.

Rump, E. E. (1968). Is there a general factor of preference for complexity? *Perception & Psychophysics*, *3*, 346–348. doi:10.3758/BF03212482

Sachs, T. (2011, June 21). 10 Bullets, #8: "ALWAYS BE KNOLLING". By Tom Sachs [Video file]. Retrieved from https://www.youtube.com/watch?v=s-CTkbHnpNQ

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*, 173–212. doi:10.1177/0022022106297299

Sherman, A., Grabowecky, M., & Suzuki, S. (2015). In the working memory of the beholder: Art appreciation is enhanced when visual complexity is compatible with working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 898–903. doi:10.1037/a0039314

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041–1053. doi:10.1037/0022-3514.84.5.1041

Stephens, K. D., & Hoffman, D. D. (2016). On visual texture preference: Can an ecological model explain why people like some textures more than others? *Perception*, *45*, 527–551. doi:10.1177/0301006616629026

Swami, V., & Furnham, A. (2012). The effects of symmetry and personality on aesthetic preferences. *Imagination, Cognition and Personality*, *32*, 41–57. doi:10.2190/IC.32.1.d

Taylor, C., & Franklin, A. (2012). The relationship between color–object associations and color preference: Further investigation of ecological valence theory. *Psychonomic Bulletin & Review*, *19*, 190–197. doi:10.3758/s13423-012-0222-1

Taylor, R. P., Spehar, B., Van Donkelaar, P., & Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock's fractals. *Frontiers in Human Neuroscience*, *5*, 60. doi:10.3389/fnhum.2011.00060

The Webbys (2015, October 12). An interview with Austin Radcliffe – founder of Webby-winning blog "Things Organized Neatly". Retrieved from http://webbyawards.com/features/an-interview-with-austin-radcliffe-founder-of-webby-winning-blog-things-organized-neatly/

Thompson, M. M., Naccarato, M. E., & Parker, K. E. (1989). Assessing cognitive need: The development of the Personal Need for Structure and the Personal Fear of Invalidity Scales. Paper presented at the Annual meeting of the Canadian Psychological Association, Halifax, Nova Scotia.

Thompson M. M., Naccarato M. E., & Parker K. E. (1992). *Measuring cognitive needs: The development and validation of the Personal Need for Structure (PNS) and Personal Fear of Invalidity (PFI) measures.* Unpublished manuscript.

Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. B. (2001). The Personal Need for Structure (PNS) and Personal Fear of Invalidity (PFI) scales: Historical perspectives, present applications and future directions. In G. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 19–39). Mahwah, NJ: Erlbaum.

Tinio, P. P. L., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, *130*, 241–250. doi:10.1016/j.actpsy.2009.01.001

Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, *10*(2), 18. doi:10.1167/10.2.18

Wagemans, J. (2011). Towards a new kind of experimental psycho-aesthetics? Reflections on the Parallellepipeda project. *i-Perception*, *2*, 648–678. doi:10.1068/i0464aap

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217. doi:10.1037/a0029333

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in Need for Cognitive Closure. *Journal of Personality and Social Psychology*, *67*, 1049–1062. doi:10.1037/0022-3514.67.6.1049

Wehrli, U. (2002). Kunst aufräumen. [Tidying up art]. Zürich: Kein & Aber AG.

Wehrli, U. (2006). Tidying up art [Video file]. Retrieved from https://www.ted.com/talks/ursus_wehrli_tidies_up_art

Wehrli, U. (2011). Die Kunst, aufzuräumen. [The art of clean up]. Zürich: Kein & Aber AG.

Wie, T., & Simko, V. (2016). corrplot: Visualization of a correlation matrix. R package (Version 0.77) [Computer software]. Retrieved from https://CRAN.R-project.org/package=corrplot

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society*, *367*, 2007–2022. doi:10.1098/rstb.2012.0098

Westphal-Fitch, G., Oh, J., & Fitch, W. T. (2013). Studying aesthetics with the method of production: Effects of context and local symmetry. *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 13–26. doi:10.1037/a0031795

Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, *23*, 165–180. doi:10.2190/B1LR-MVF3-F36X-XR64

Wundt, W. (1874). *Grundzüge der physiologischen psychologie* [Fundamentals of physiological psychology]. Leipzig, Germany: Wilhelm Engelmann.

Appendix A

Stimulus Pair Selection

For my project I started out with searching for images of neatly organized compositions that were available online. Based on the series of images that I collected and the literature reviewed above, I sought for stimulus dimensions related to the balance between order and complexity that could play a role in the aesthetic preference for certain neatly organized images. Each of these qualitatively determined factors will be discussed below. Then I selected pairs of stimuli that differed in order and/or complexity on one or maximally two of these perceptual stimulus dimensions, to be able to investigate the importance of order and complexity on each of these stimulus dimensions in determining aesthetic preference. All image pairs are presented in full resolution and size on http://gestaltrevision.be/tests/ton1/allpics.php.

a

b

Figure A1. Example of a stimulus pair testing low versus medium color complexity (a) and of a stimulus pair testing medium versus high color complexity (b).

Color order and complexity. One of the factors assumed to play a role in the aesthetic preference for specific images of neatly organized compositions is the amount of variety in color present in the image. Within this category two types of pairs were distinguished: pairs comparing low and medium color complexity (e.g., Figure A1.a) and pairs comparing medium and high color complexity (e.g., Figure A1.b). The first type of pairs compared images in which only one main color is present with images in which different (often compatible) colors are present. The second type of pairs compared images in which different compatible colors are present with images in which different more contrasting colors are present. The differences between the images in this second type of pairs are less unambiguous than the differences between the images in the first type of pairs, partially because also difference in color order come to play here. That is, some of the images in this second type of pairs do contain a color gradient (i.e., systematic alteration in color; e.g., COL_18) or another type of systematic ordering in color (e.g., COL_21), which may mask pure differences in color complexity between the images.

a b

Figure A2. Example of a stimulus pair testing texture complexity (a) and a stimulus pair testing global form and contour complexity (b).

Textural order and complexity. With texture I refer to the physical composition or structure of the different elements in the image (i.e., how the elements are placed in relation towards each other). Whereas in some of the pairs images with a regular texture and those with a more irregular texture pattern are compared (e.g., Figure A2.a), other pairs were selected to tap in differences in aesthetic preferences for different regular patterns (e.g., TEX_2 comparing horizontal and diagonal texture based on color differences; TEX_10 all elements on a fixed distance of each other versus no distance between the elements; TEX_17 subgroups of elements versus gradient, etc.). In some of the image pairs color plays a role in defining the difference in texture between the images (e.g., TEX_2).

Global form and contour order and complexity. Some of the image pairs were selected to differ in the complexity of the global form and contour arising from the combination of the different elements in the image (e.g., Figure A2.b). In some of the images the contour of the collection of elements in the image is emphasized, which could add a dimension of order or coherence to these stimuli (e.g. right image in CON_6).

a b

Figure A3. Example of a stimulus pair testing a difference in number of objects (a) and a stimulus pair testing a difference in perspective (b).

Number of objects. The number of elements in a stimulus or image is a well-known aspect of complexity that also got its place in this study (e.g., Figure A3.a).

Perspective. Whereas most images of neatly organized compositions are taken from above, some images are taken from a different – more complex looking – perspective (e.g., Figure A3.b).

Figure A4. Example of a stimulus pair testing a difference in type of objects.

Type of objects. Whether images contain very different types of objects or only very similar objects (i.e., on a basic categorical level) can also influence complexity and consequently aesthetic preferences (e.g., Figure A4).

a

b

c

d

Figure A5. Example of a stimulus pair testing the factor combination color and configuration (a), a stimulus pair testing the factor combination color and type of objects (b), a stimulus pair testing the factor combination color and texture (c), and a stimulus pair testing the factor combination texture and number of objects (d).

Combinations of factors. Some of the selected image pairs differ in two of the mentioned dimensions simultaneously (see Figure A5.a–A5.d). When one of the images in a pair is assumed to be more ordered or complex on both dimensions, this combination of factors can lead to clearer differences in complexity or order (e.g., Figure A5.a and A5.d). When one of the images is assumed to be more ordered or complex on one dimension and the other image is assumed to be more ordered or complex on the other dimension, this pair can give insight in which factor has more weight in the judgment of order and complexity and in aesthetic preference judgments (e.g., Figure A5.b and A5.c).

Other considerations concerning selection of stimulus pairs and dimensions. For factors or dimensions that were not systematically varied or taken into account, it was attempted to make the selected sample of images representative for the collection of images available online. Furthermore, images showing an extreme (non-)ordering or complexity on these non-selected dimensions were removed from the selection. Some of these non-selected factors were the shape of the objects, length or size of the objects, and prominence of the background.

Examples of images containing important conceptual aspects that were not taken into account were: (a) images in which the meaning of the presented objects was lost after ordering (e.g., Chinese letters); (b) images in which one object is analyzed in its parts (in contrast to ‘arrange’ images, in which different objects are organized in a collection); (c) images containing a conceptual ordering (e.g., objects organized in a form with the contour of the United States of America); and (d) images of non-real-life objects (e.g., ordered paintings). We decided to only show organized images (and not combinations of pre- and after-ordering images) as the starting non-ordered counterpart was not available for most ordered images.

Appendix B

Coding Scheme of the Stimulus Pairs

Below the original coding scheme of the stimulus pairs is reported. Except for the color dimension, all codings are very subjective and arbitrary. Additionally, it was hard to make a clear distinction between differences in complexity (i.e., amount of variation on a dimension) and differences in order (i.e., organization or regularity of the variation on that dimension). This results in an often overlapping coding for differences in order and complexity (i.e., 1 for complex, – 1 for order). Sometimes, two codings or a question mark are shown to indicate that the coding was not straightforward for that difference in the pair.

A score of 1 indicates that id1 is assumed to be more ordered or complex than id0. A score of – 1 indicates that id1 is assumed to be less ordered or complex than id0. A score of 0 indicates that there is no evident difference in order or complexity on that dimension between the images in the pair.

The names of the images that were used twice are in italic. Besides the varied dimensions (i.e., order and complexity differences in color, texture, global form/contour, number of objects, type of objects, and perspective), the stimulus pairs were also coded for possible differences in the prominence of the background as well as differences in the recognizability of the objects.

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
COL_1	red13-bewerkt	colours6-bewerkt	1	1	0	0	0	0	1	-1	0	1	0	0	0
COL_2	pretzels1-bewerkt	food48-bewerkt	1	1	0	0	0	0	1	-1	0	0	0	1	0
COL_3	kitchen2	kitchen3	1	1	0	0	0	0	0	0	0	-1	0	0	0
COL_4	leaves-12-bewerkt	*leaves-11-bewerkt*	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_5	red18-bewerkt	stuff31-bewerkt	1	1	0	0	0	0	1	?	?	-1	0	1	1
COL_6	*blue plastic2-bewerkt2*	*plastic3-bewerkt*	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_7	leafs13-bewerkt	*leafs18-bewerkt*	1	1	0	0	0	0	1	-1	1	0	0	1	0
COL_8	*Jo44*	*Erasers1-other version-bewerkt*	1	1	0	0	0	0	0	0	0	1	0	0	0
COL_9	bottles6	bottles7	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_10	cassette5-bewerkt	cassette1	1	1	0	0	0	0	0	0	0	0	0	-1	0
COL_11	flowers-7-bewerkt	flowers-5	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_12	Jo53-bewerkt	Jo51-bewerkt	1	1	0	0	0	0	0	0	0	0	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
COL_13	Jo54-bewerkt	Jo49-bewerkt	1	1	0	0	0	0	0	0	0	1	0	0	0
COL_14	hands1-bewerkt	hands2-bewerkt	1	1	0	0	0	0	0	0	0	0	0	1	0
COL_15	colours5-bewerkt	colours4-bewerkt	1	1	-1	0	0	0	0	0	0	0	0	1	0
COL_16	colours7-bewerkt	colours13-bewerkt	1	0	1	0	0	0	0	0	0	0	0	0	0
COL_17	pens_5-bewerkt	Jules2-bewerkt	1	1	1	0	0	0	0	0	0	0	0	0	0
COL_18	*flowers-12*	Jules7-bewerkt	1	1	1	0	0	0	1	1	1	0	0	0	0
COL_19	colours-5-bewerkt	colours-4-bewerkt	1	1	-1	0	0	0	0	0	0	0	0	-1	0
COL_20	Jules22-bewerkt	Jules6-R-bewerkt	1	1	0	0	0	0	0	0	0	0	0	0	-1
COL_21	dice_3-bewerkt2	*dice1-bewerkt*	1	1	-1	0	0	0	0	0	0	0	0	-1	0
COL_22	*plastic3-bewerkt*	Beach treasures1-bewerkt	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_23	food37-bewerkt	Jules27-bewerkt	1	1	0	0	0	0	0	0	0	0	0	0	0
COL_24	eggs1(2)	eggs3	1	1	-1	0	0	0	0	0	0	0	0	1	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
COL_25	buttons-1-bewerkt	buttons7-bewerkt	1	1	-1	0	0	0	0	0	0	0	0	1	0
TEX_1	candy12	candy13	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_2	Jules12-bewerkt	Jules11-bewerkt	0	0	0	0	0	0	1	1	0	0	0	0	0
TEX_3	leaves-1-bewerkt	*leaves-6-bewerkt*	0	0	0	0	0	0	1	1	0	0	0	0	0
TEX_4	black6-bewerkt	black9	0	0	0	0	0	0	1	1	0	0	0	0	0
TEX_5	shoes17-bewerkt	*shoes4-bewerkt*	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_6	flowers-6-bewerkt	*flowers-12*	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_7	fruit14-bewerkt	*fruit-2-bewerkt2*	0	0	0	0	0	0	1	1	1	0	0	0	0
TEX_8	tomatoes4-bewerkt3	*fruit-2-bewerkt2*	0	0	0	0	0	0	1	-1	1	0	0	0	0
TEX_9	candy16	candy10	0	0	0	0	0	0	1	1	-1	0	0	1/-1	0
TEX_10	buttons2-bewerkt	buttons6-bewerkt	1	-1	0	0	0	0	1	1	-1	0	0	0	0
TEX_11	flowers-4-bewerkt	flowers-2-bewerkt	0	0	0	0	0	0	1	1	-1	0	0	1	0
TEX_12	candy14-bewerkt	candy15-bewerkt	1	-1	0	0	0	0	1	1	0	0	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
TEX_13	medicine5-bewerkt	candy20-bewerkt	0	0	0	0	0	0	1	0	-1	0	0	-1	0
TEX_14	food16	tomatoes2	1	-1	0	0	0	0	1	1	0	0	0	0	0
TEX_15	shoes9	*shoes4-bewerkt*	1	-1	0	0	0	0	1	1	-1	0	0	0	0
TEX_16	stones-2-bewerkt	stones13-bewerkt	0	0	0	0	0	0	1	1	1	0	0	1	0
TEX_17	Jo42-bewerkt	*Erasers1-other version-bewerkt*	0	0	0	0	0	0	1	0	-1	0	0	0	0
TEX_18	leafs3-bewerkt	*leaves-16-bewerkt*	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_19	citrons-1-bewerkt	citrons-2-bewerkt	0	0	0	0	0	0	1	1	-1	0	0	1	0
TEX_20	cars3-bewerkt	cars4-bewerkt	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_21	bottles12-bewerkt	Jules23-bewerkt	0	0	0	0	0	0	1	?	?	0	0	0	0
TEX_22	*leaves-16-bewerkt*	leaves-2-bewerkt	1	1	0	0	0	0	1	1	-1	0	0	0	0
TEX_23	bottles2	bottles1	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_24	brush7	brush8	0	0	0	0	0	0	1	1	-1	0	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
TEX_25	clips4-bewerkt	clips2	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_26	corkshrews1	corkshrews2	0	0	0	0	0	0	1	1	-1	1	0	0	0
TEX_27	fashion9	fashion7	0	0	0	0	0	0	1	1	0	0	0	0	0
TEX_28	knifes2	knifes8	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_29	pens26	pens_1	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_30	*sewing2-bewerkt2*	sewing4-bewerkt	1	-1	-1	0	0	0	1	1?	0	0	0	0	0
TEX_31	spoons1-bewerkt2	*spoons-1-bewerkt*	0	0	0	0	0	0	1	1	-1	0	0	1	0
TEX_32	*spoons-1-bewerkt*	spoons16	0	0	0	0	0	0	1	1	0	0	0	1	0
TEX_33	colours-6	Jules15	0	0	0	0	0	0	1	1?	-1?	0	1	0	0
TEX_34	red-1-bewerkt	red15-bewerkt	0	0	0	0	0	0	1	1	-1	0	0	0	0
TEX_35	flowers10	white2	0	0	0	0	0	0	1	1/-1?	1	0	0	0	0
TEX_36	flowers-9	yellow2	0	0	0	0	0	0	1	1/-1?	1	0	0	0	0
TEX_37	flowers-13	pink2	0	0	0	0	0	0	1	1/-1?	1	0	0	0	0
TEX_38	flowers-11	red2	0	0	0	0	0	0	1	1/-1?	1	0	0	0	0
TEX_39	candy11	candy4	0	0	0	0	0	0	1	1	-1	0	1	0	0
CON_1	bat1	bat2	0	0	0	1	1	-1	0	0	0	0	0	0	0
CON_2	blue14-bewerkt	blue2-bewerkt	0	0	0	1	1	-1	0	0	0	0	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
CON_3	black28-bewerkt	black27-bewerkt2	0	0	0	1	1	-1	0	0	0	1	0	1	0
CON_4	green6	green7	0	0	0	1	1	-1	0	0	0	0	0	0	0
CON_5	stones-1-bewerkt	*leaves-6-bewerkt*	0	0	0	1	1	-1	0	0	0	0	1	0	0
CON_6	paint2	Jules26	0	0	0	1	0	1	0	0	0	0	0	0	0
CON_7	notes6-bewerkt	notes3-bewerkt	1	-1	0	1	1	-1	1	?	?	0	0	1	0
CON_8	keys6-bewerkt1	keys7-bewerkt2	0	0	0	1	1	-1	0	0	0	0	0	0	0
CON_9	glass-1-bewerkt	zip1	0	0	0	1	1	0	0	0	0	0	0	0	0
CON_10	black21-bewerkt	black_1-bewerkt	0	0	0	1	0	1	0	0	0	0	0	0	0
PRP_1	perspcontr2-bewerkt	persp2	0	0	0	1	1	-1	0	0	0	0	0	0	0
PRP_2	*pink-1-bewerkt*	persp1	0	0	0	1	1	-1	0	0	0	0	0	0	0
PRP_3	perscontr3	persp3	0	0	0	1	1	-1	0	0	0	0	0	0	0
NUM_1	glasses1	glasses2	1	1	0	0	0	0	0	0	0	1	0	1	0
NUM_2	*pink-1-bewerkt*	pink6-bewerkt	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_3	black7	black12	0	0	0	0	0	0	0	0	0	1	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
NUM_4	orange4-bewerkt2	orange3-bewerkt	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_5	blue plastic3-better-bewerkt	*blue plastic2-bewerkt2*	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_6	bread3	bread2-bewerkt	0	0	0	0	0	0	0	0	0	1	0	1	0
NUM_7	blue16-bewerkt	blue15-bewerkt	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_8	digital12	digital10	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_9	digital20	digital18	0	0	0	0	0	0	0	0	0	1	0	0	0
NUM_10	red16-bewerkt	red14-bewerkt	0	0	0	0	0	0	0	0	0	1	0	0	0
TYP_1	*leafs18-bewerkt*	leaves-10-bewerkt	1	-1	-1	0	0	0	0	0	0	0	1	0	0
TYP_2	*leaves-11-bewerkt*	leaves-13-bewerkt	0	0	0	0	0	0	0	0	0	0	1	0	0
CL+CN_1	yellow3	stuff30	1	1	0	1	1	-1	0	0	0	0	0	0	0
CL+CN_2	leafs8	leafs12	1	1	0	1	0	1	0	0	0	0	0	1	0
CL+TP_1	*Jo44*	*buttons3*	1	1	0	0	0	0	1	-1	1	0	1	0	0
CL+TX_1	*sewing2-bewerkt2*	needles-4-bewerkt	1	-1	0	0	0	0	1	1	-1	1	1	0	0
CL+TX_2	*buttons3*	Jules16	1	1	-1	0	0	0	1	1	-1	1	0	0	0

			COLOR			FORM/CONTOUR			TEXTURE			NUMBER OF OBJECTS	TYPE OF OBJECTS	BACK-GROUND	RECOGNIZ-ABILITY
PAIR NAME	ID0	ID1	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	COMPLEXITY	ORDER	DIFF?	DIFF?	DIFF?	DIFF?
CL+TX_3	flowers8	food38	1	-1	0	0	0	0	1	1	-1	0	1	1	0
CL+TX_4	stones-4-bewerkt	Jules24-bewerkt	1	1	1	0	0	0	1	1	?	0	0	1	0
CL+TX_5	*dice1-bewerkt*	dice2-bewerkt	1	1	1	0	0	0	1	1	-1	1	0	0	0
TX+NM_1	gums1-bewerkt	gums2-bewerkt	0	0	0	0	0	0	1	1	-1	1	0	0	0
TX+NM_2	clips5-bewerkt	clips-1-bewerkt	0	0	0	0	0	0	1	1	-1	1	0	0	0
TX+NM_3	brush6-bewerkt	brush14-bewerkt	0	0	0	0	0	0	1	1	0	1	0	0	0

Appendix C

Supplementary Tables

Table S1

Participant Characteristics Part 1

	Dutch version		English version		Total	
Measure	N	%	N	%	N	%
	359	100.00	56	100.00	415	100.00
Gender						
Female	238	66.30	35	62.50	273	65.78
Male	121	33.70	21	37.50	142	34.22
Other	0	0.00	0	0.00	0	0.00
Age (Binned)						
16-19	25	6.96	2	3.57	27	6.51
20-24	68	18.94	11	19.64	79	19.04
25-29	29	8.08	16	28.57	45	10.84
30-34	32	8.91	8	14.29	40	9.64
35-39	34	9.47	4	7.14	38	9.16
40-44	23	6.41	4	7.14	27	6.51
45-49	29	8.08	4	7.14	33	7.95
50-54	36	10.03	3	5.36	39	9.40
55-59	32	8.91	1	1.79	33	7.95
60-64	30	8.36	1	1.79	31	7.47
65-69	12	3.34	1	1.79	13	3.13
70-74	5	1.39	0	0.00	5	1.20
75-79	4	1.11	1	1.79	5	1.20
Mother Tongue (Grouped)						
Dutch	349	97.21	3	5.36	352	84.82
Romance	2	0.56	13	23.21	15	3.61
English	1	0.28	12	21.43	13	3.13
Slavic	2	0.56	9	16.07	11	2.65
German	4	1.11	6	10.71	10	2.41
Greek	0	0.00	5	8.93	5	1.20
Other	1	0.28	8	14.29	9	2.17

	Dutch version		English version		Total	
Measure	N	%	N	%	N	%
Highest Level of Education						
Elementary	1	0.28	0	0.00	1	0.24
Part of high school (no diploma)	25	6.96	0	0.00	25	6.02
High school	106	29.53	8	14.29	114	27.47
College	119	33.15	2	3.57	121	29.16
University	108	30.08	46	82.14	154	37.11

Note. Romance languages include French, Italian, Romanian, Portuguese (Brazilian), and Spanish.

Slavic languages include Bulgarian, Czech, Polish, Russian, Serbian, and Slovenian.

Other languages include Arabic, Chinese, Estonian, Hebrew, Norwegian, Persian, and Turkish.

Table S2

Participant Characteristics Part 2

	Dutch version		English version		Total	
Measure	N	%	N	%	N	%
	76	100.00	8	100.00	84	100.00
Gender						
Female	49	64.47	7	87.50	273	65.78
Male	27	35.53	1	12.50	142	34.22
Other	0	0.00	0	0.00	0	0.00
Age (Binned)						
16-19	0	0.00	0	0.00	0	0.00
20-24	20	26.32	2	25.00	22	26.19
25-29	2	2.63	2	25.00	4	4.76
30-34	2	2.63	2	25.00	4	4.76
35-39	5	6.58	0	0.00	5	5.95
40-44	5	6.58	1	12.50	6	7.14
45-49	10	13.16	0	0.00	10	11.90
50-54	7	9.21	0	0.00	7	8.33
55-59	7	9.21	0	0.00	7	8.33
60-64	12	15.79	0	0.00	12	14.29
65-69	4	5.26	1	12.50	5	5.95
70-74	1	1.32	0	0.00	1	1.19
75-79	1	1.32	0	0.00	1	1.19

	Dutch version		English version		Total	
Measure	N	%	N	%	N	%
Mother Tongue (Grouped)						
Dutch	74	97.37	0	0.00	74	88.10
Romance	1	1.32	2	25.00	3	3.57
English	0	0.00	4	50.00	4	4.76
Slavic	0	0.00	1	12.50	1	1.19
German	1	1.32	0	0.00	1	1.19
Greek	0	0.00	0	0.00	0	0.00
Other	0	0.00	1	12.5	1	1.19
Highest Level of Education						
Elementary	0	0.00	0	0.00	0	0.00
Part of high school (no diploma)	2	2.63	0	0.00	2	2.38
High school	23	30.26	1	12.50	24	28.57
College	28	36.84	1	12.50	29	34.52
University	23	30.26	6	75.00	29	34.52

Note. Romance languages include French, Italian, and Portuguese (Brazilian). Slavic languages include Serbian. Other languages include Hebrew.

Appendix D

Supplementary Figures

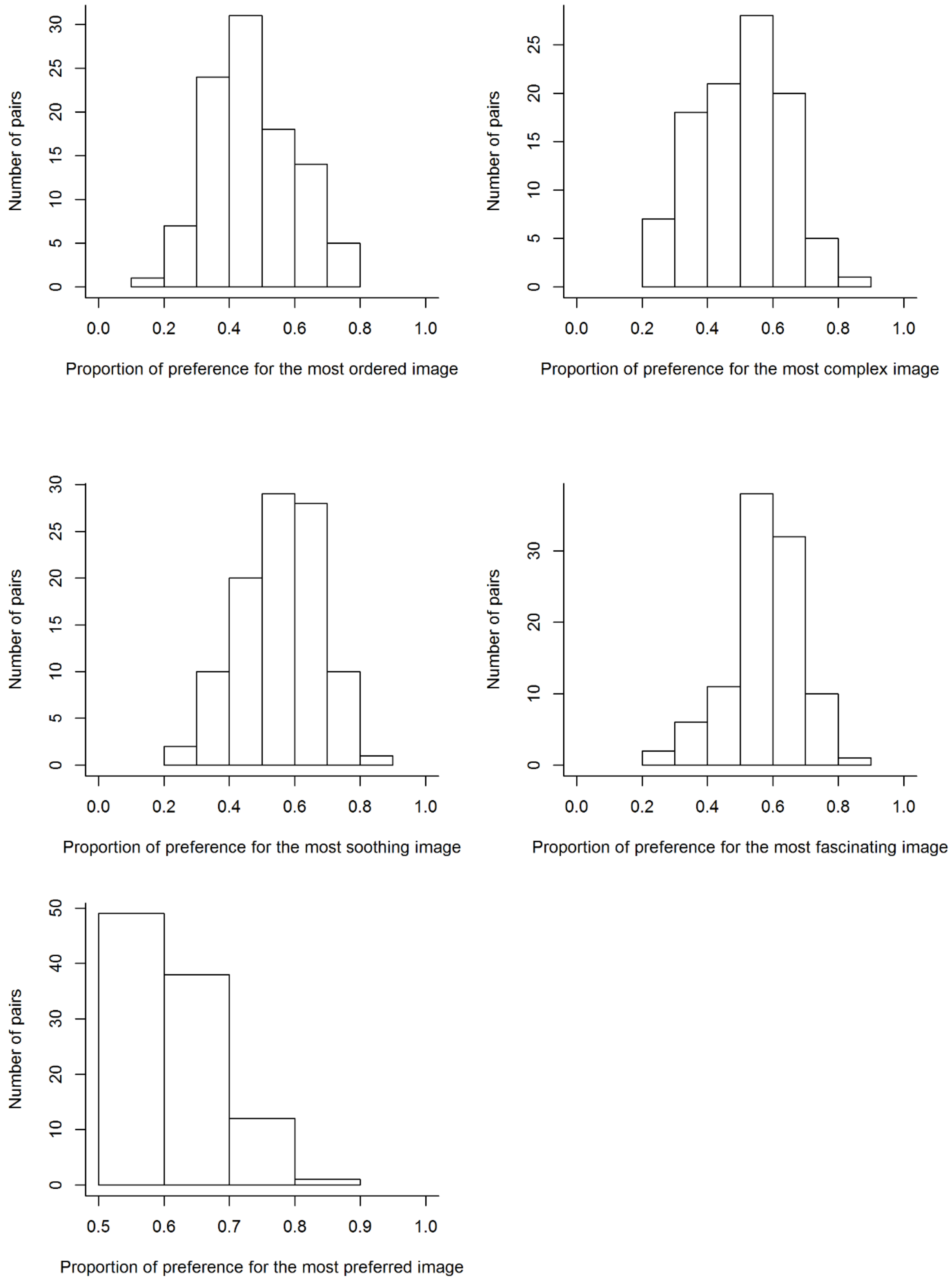


Figure S1. Histograms of the proportion of participants that preferred the most ordered, complex, soothing, fascinating, or on average preferred image, across image pairs (*N* = 100).

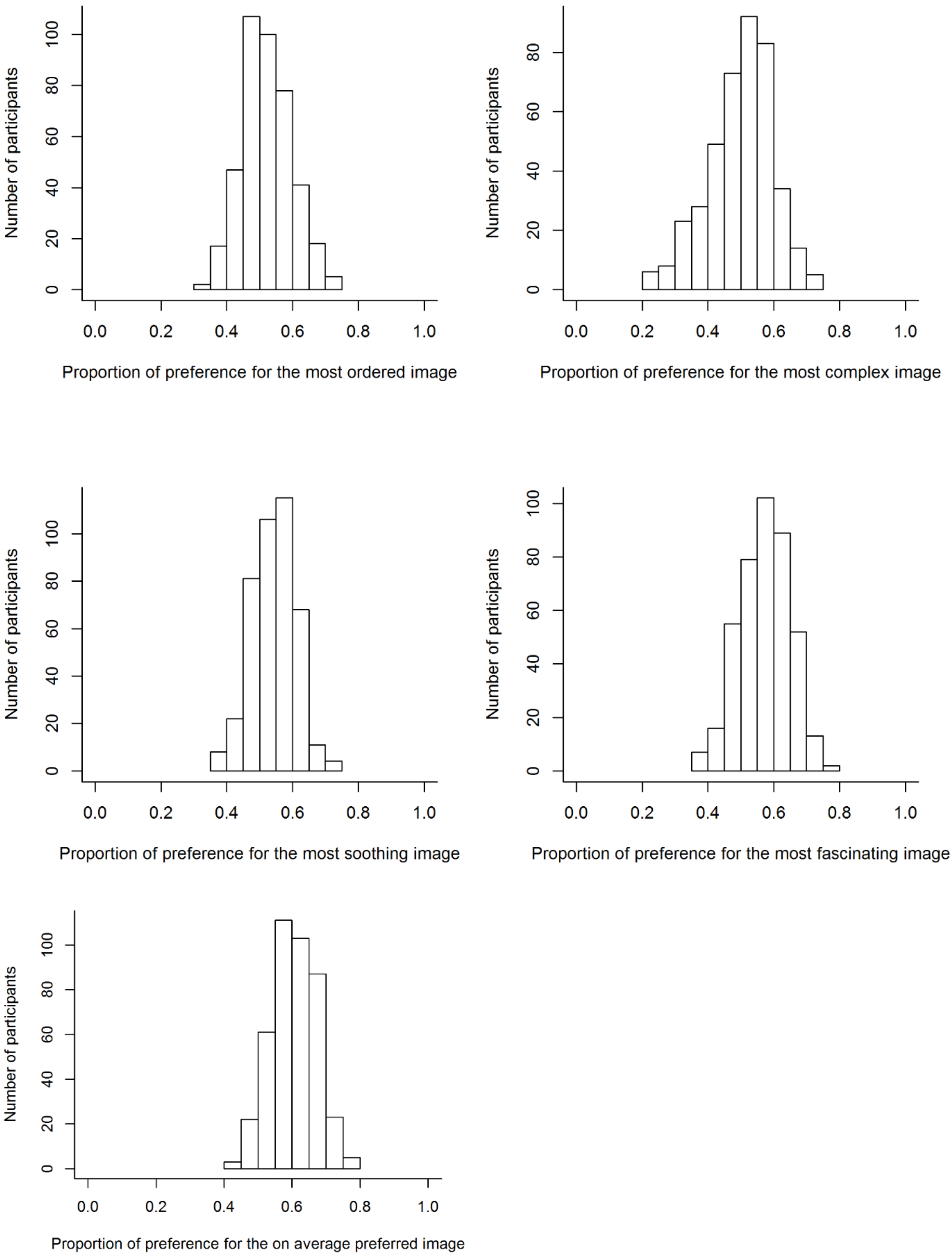
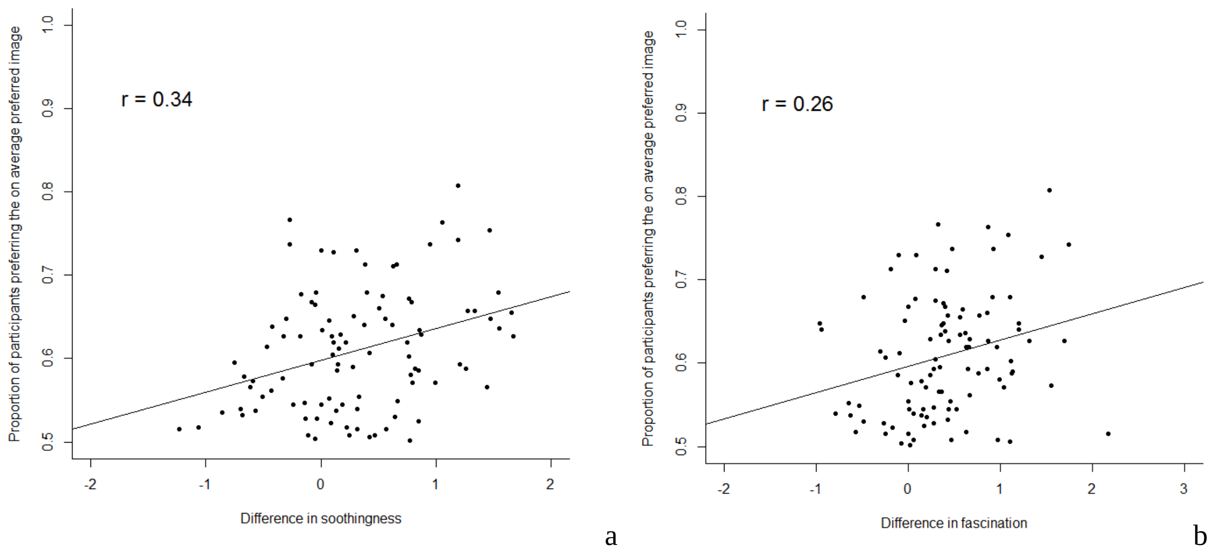


Figure S2. Histograms of the proportion of pairs in which the most ordered, complex, soothing, and fascinating, or on average preferred image was preferred, across participants (*N* = 415).



a b

Figure S3. Scatter plots showing the relation between the proportion of participants preferring the on average preferred image and the difference in soothingness (a) and fascination (b) for each of the image pairs (standardized scores, based on preference coding).

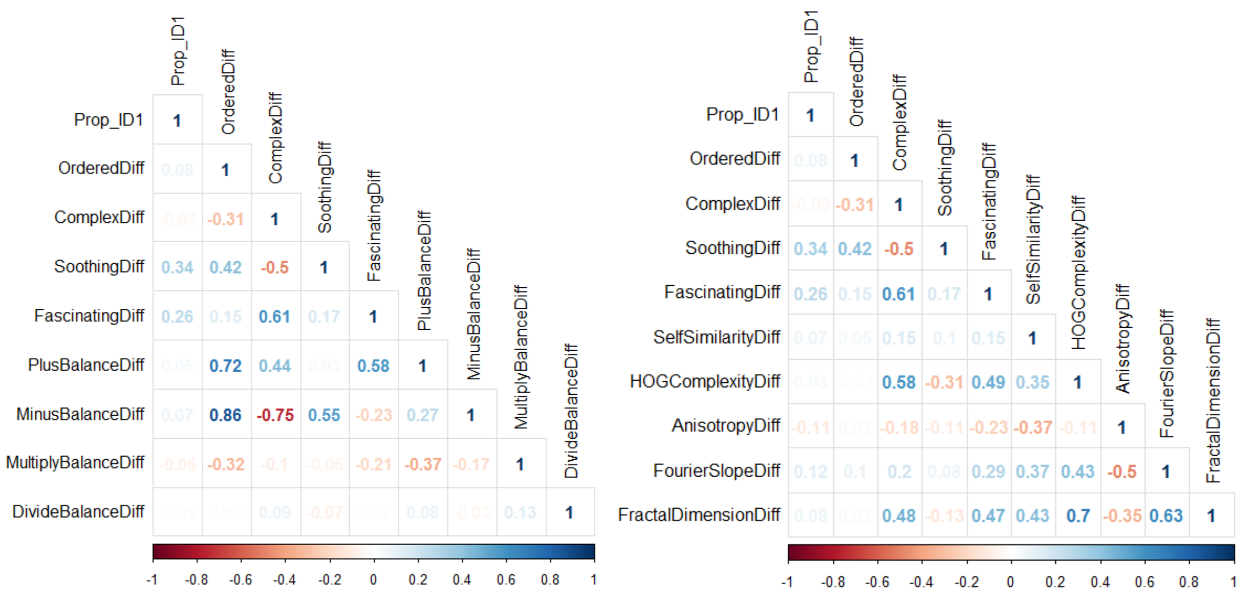


Figure S4. Pearson product-moment correlations between different image pair measures (standardized scores, based on preference coding).

Note. Prop_ID1 indicates the proportion of preference for the on average preferred image in each pair.

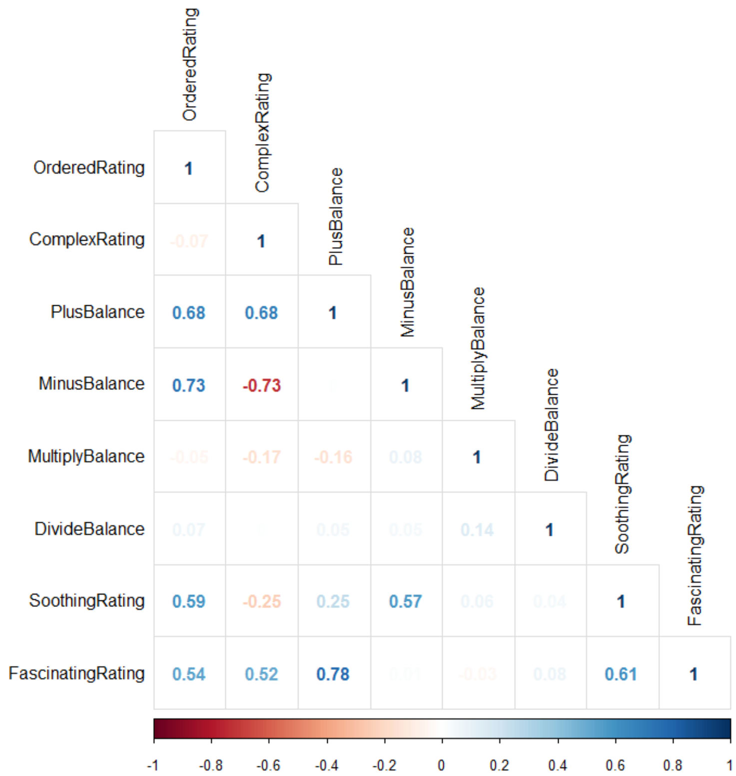


Figure S5. Correlations between average rating scores per image and balance measures between order and complexity (*N* = 184; all standardized scores).

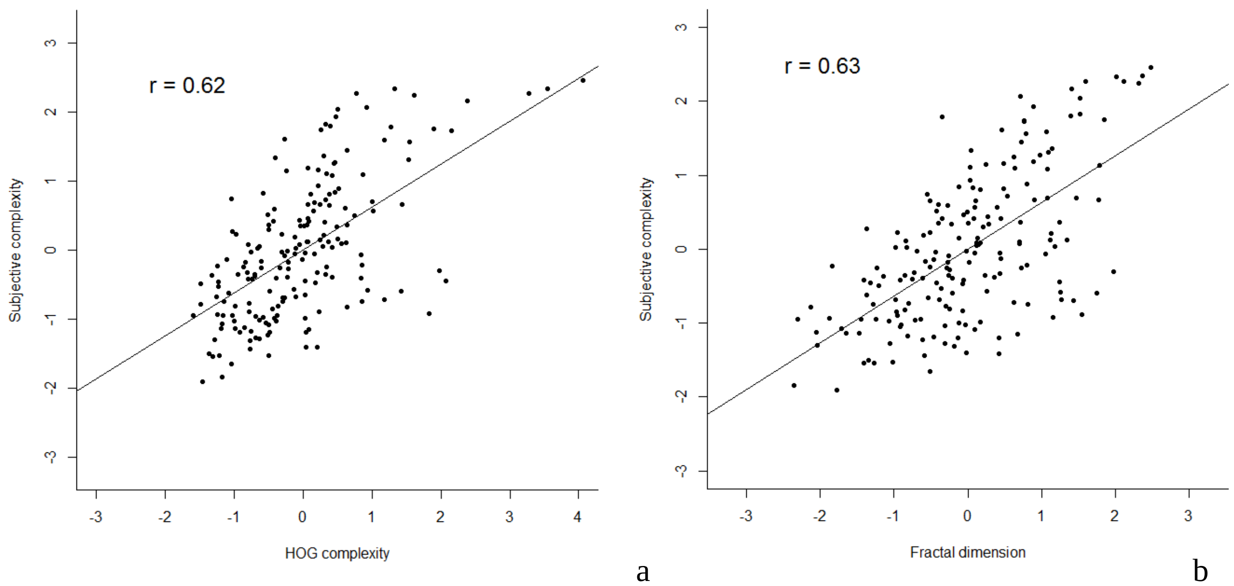


Figure S6. Scatter plots showing the relation of subjective complexity with HOG complexity (a) and fractal dimension (b) for each of the images (*N* = 184; all standardized scores).

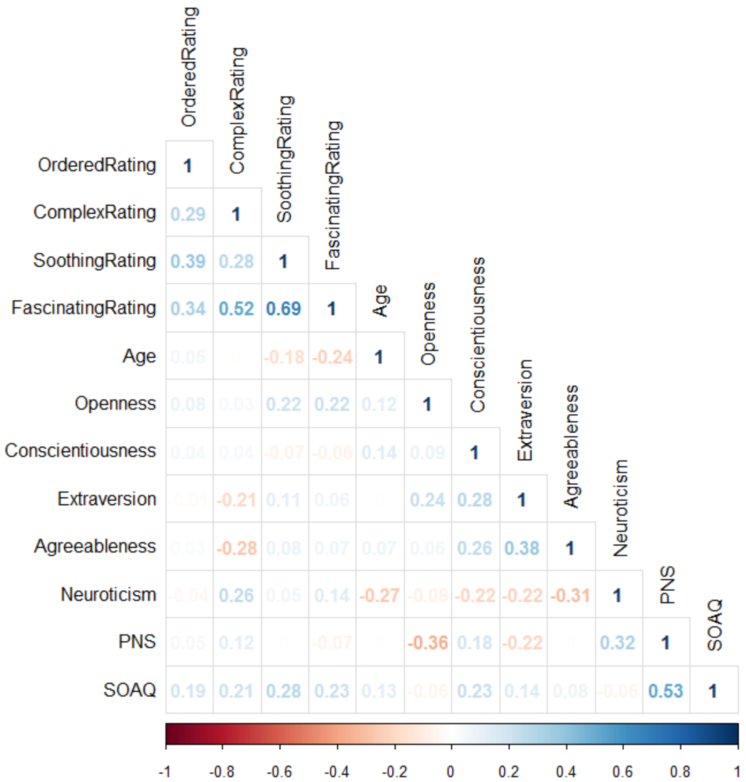


Figure S7. Correlations between average rating scores per participant and other participant measures (N = 84; unstandardized scores).