Differentiating the visual aesthetics of the sublime and the beautiful; Selective effects of stimulus size, height, and color on sublimity and beauty ratings in photographs

Young-Jin Hur, Christian Hallam-Evans, Yvette Garfen, Adam Baiza, Tabitha Backhouse Spriggs, Margareta-Theodora Mircea, Oscar Nagy, Emily Pye, & Chris McManus

London College of Fashion & University College London

Author Note

Young-Jin Hur, Fashion Business School, London College of Fashion, University of the Arts London, UK & Department of Clinical, Educational and Health Psychology, University College London, UK.

Christian Hallam-Evans, Department of Experimental Psychology, University College London, UK

Yvette Garfen, Department of Experimental Psychology, University College London, UK

Adam Baiza, Department of Experimental Psychology, University College London, UK

Tabitha Backhouse Spriggs, Department of Experimental Psychology, University College London, UK

Margareta-Theodora Mircea, Department of Experimental Psychology, University College London, UK

Oscar Nagy, Department of Experimental Psychology, University College London, UK

Emily Pye, Department of Experimental Psychology, University College London, UK

I. Chris McManus, Department of Clinical, Educational and Health Psychology & Research Department for Medical Education (RDME), University College London, UK.

Corresponding Author: Young-Jin Hur, Email: yhur1885@gmail.com. Tel: +44 (0)7715448313.

Abstract

Despite the philosophical literature concerning the sublime and the beautiful, dating back to Burke (1759/2008), there is still limited empirical evidence regarding the visual aesthetics of sublimity and beauty. The present paper asks whether the manner in which photographs are presented can alter the perception of the sublimity and beauty ratings of these photographs. In a set of studies, it is reported that the increase of presentation size increases sublimity more than beauty (Study 1) and that this is mainly driven by the effects of visual angle (Study 3). While increasing presentation height affects both sublimity and beauty positively and in similar degrees (Study 1), the presence of color (vs. black and white [monochrome]) is predominantly related to judgments of beauty (Study 2). Brightness and contrast levels affected neither sublimity nor beauty (Study 3). An important methodological point is that all inferential statistics use linear mixed models, which treat both participants and stimuli as random effects. In addition, each participant receives different random subsets of stimuli, increasing the size of the stimulus set. Overall, the analyses incorporate 233 photographs and 245 participants in total which allows the generalizability of findings. Sublimity and beauty respond differentially to different presentational cues, which demonstrates the importance of simultaneously considering sublimity and beauty in empirical studies on aesthetic judgments.

Keywords: sublime, beauty, size, empirical aesthetics, awe

Introduction

The sublime remains one of the most enduring topics in philosophical aesthetics, dating back to Longinus' *On the Sublime* (1st century AD). In various philosophical descriptions, particularly that of Edmund Burke's *A Philosophical Enquiry in the Origins of our Ideas of the Sublime and Beautiful* of 1759 (Burke, 1759/2008), the sublime has been commonly characterized as a delight that contains elements of power and reverence, most often associated with grand nature (e.g. the Alps & the sea).

An important assumption in these philosophical descriptions of the sublime is that the overwhelming, often shocking qualities of sublimity may be an experience that contrasts to the more controlled pleasures of beauty. Importantly, the experiences of sublimity and beauty have been argued to be selectively affected by certain physical characteristics. It is often argued, for instance, that sublimity may be elicited through physically large objects while beauty may be elicited through physically small objects (Ashfield and de Bolla, 1996; Monk, 1935; Nicolson, 1959).

While the sublime has received some scientific interest of late (e.g. Hur Gerger, Leder, & McManus, 2020), no research to date has systematically explored the question of whether certain visual manipulations selectively influence sublimity and beauty. Specifically, despite the philosophical tradition of associating sublimity with certain visual characteristics in objects (e.g. largeness), works have also argued for different or opposite viewpoints, arguing, for example, how experiences of sublimity as well as beauty may manifest in small objects (e.g. Pelowski, Hur… McManus, 2021). In light of the historical yet contradictory literature, the present paper presents a series of experiments where sublimity and beauty ratings toward photographs can be selectively induced by the manipulations of photograph size, height, and color.

Sublimity versus Beauty

At the extreme, sublimity and beauty may present opposing psychologies based on different physical elicitors. In the quotation below, from Burke's *A Philosophical Enquiry into the Origins of our Ideas of the Sublime and Beautiful* (1759/2008), is shown a typical view that sublimity and beauty are founded on different psychological bases that are triggered by different physical causes.

> "For sublime objects are vast in their dimensions, beautiful ones comparatively small; beauty should be smooth and polished; the great, rugged and negligent: beauty should shun the right line, yet deviate from it insensibly; the great in many cases loves the right line; and when it deviates, it often makes a strong deviation: beauty should not be obscure; the great ought to be dark and gloomy: beauty should be light and delicate; the great ought to be solid, and even massive. They are indeed ideas of a very different nature..." (p.123)

This historical viewpoint on sublimity and beauty notwithstanding, there are only a handful of empirical works that have assessed sublimity's relationship with beauty. Ishizu and Zeki (2014) notably explored the different neural mechanisms underlying sublimity and beauty judgments using fMRI measures. For example, where the body of the caudate nucleus (associated with the experiences of beauty and romantic love; Ishizu & Zeki, 2011) and the medial orbitofrontal cortex (associated with the experiences of beauty, pleasure, reward, and value; Ishizu & Zeki, 2011) were activated with visual beauty but not sublimity, the head of the caudate (associated with cognitive and emotional functions, but especially associated with the former; Ishizu & Zeki, 2014) was uniquely activated with visual sublimity but not beauty.

Furthermore, Hur et al. (2020) reported that ratings of sublimity and beauty in photography were associated with different subjective emotional states. Specifically, replicated in two separate studies in Hur et al. (2020), beauty was correlated negatively with fear, positively with happiness, and not correlated with arousal. On the other hand, sublimity was positively correlated with fear, not correlated with happiness, and positively correlated with arousal. When partial correlations were further run to examine each aesthetic experience's emotional profile while controlling for the other aesthetic experience, sublimity and beauty revealed diametrically opposing emotional profiles; beauty was correlated with low fear, high happiness, and low arousal, whereas sublimity was correlated with high fear, low happiness, and high arousal.

Relatedly, research by Marković (2010) demonstrated that when participants were asked to rate a set of paintings with a set of 22 affective descriptors, factor analysis of the descriptors revealed two main factors. Factor 1 included typical beauty-related descriptors (e.g. lovely, charming, appealing, etc.; see also Marković [2014], where pleasure was the best predictor for the experience of

beauty) whereas Factor 2 included typical sublimity-related descriptors (e.g. exceptional, profound, awing, unspeakable, etc.). The correlation between the two factors was fairly low ($r = 0.11$). This limited empirical literature, thus, suggests that sublimity and beauty may operate on distinct psychologies (Hur, 2020; Marković, 2012).

It must be noted that while sublimity and beauty are increasingly studied in psychology, they were rarely assessed simultaneously within a single experimental study (e.g. Eskine, Kacinik, & Prinz, 2012). Methodologically, the measure of either sublimity *or* beauty rather than the measure of both sublimity *and* beauty can pose interpretational challenges. When measuring either one of the two dimensions, the measure of sublimity, for instance, may reflect three possibilities: beauty instead of sublimity, both beauty and sublimity, or sublimity instead of beauty. Conversely, the simultaneous measurements of sublimity and beauty may enable simultaneous and differential assessments of the two concepts.

Thus, except for some works (e.g. Marković, 2010), sublimity in psychology is often depicted as an umbrella term to denote a wide range of non-specific positive aesthetic phenomena, many of them resembling characteristics of what may also be called "beautiful" (Hur & McManus, 2017; Hur, 2020; Pelowski et al., 2021); in other words, the sublimity-beauty relationship that historical texts repeatedly suggest is fundamentally underexplored. Therefore, the simultaneous measurement of both sublimity and beauty in empirical works may be a crucial step in better understanding not only the relationship between sublimity and beauty but also sublimity in the context of beauty and vice versa. Assuming that sublimity and beauty may indeed be operant on separate psychologies and elicited by separate physical realities, below is a review of some object features that sublimity and beauty experiences may be distinctly associated with.

Size and Height

Size and height have been considered as the two key characteristics of objects associated with sublimity (Costelloe, 2012). In the English language, sublimity often finds associations with physical features of largeness and height. In the *Oxford English Dictionary*, for instance, the sublime is

characterized as “rising to a great height; lofty, towering”, “uplifting”, “grand and elevated”, and “vastness.”[1]

These characteristics can often be detected in descriptions of the sublime in the natural world. For example, 18th-century British writer Addison described the sublime as experiences deriving from “the Prospects of an open Champaign Country, a vast Desert, a huge Heap of Mountains, high rocks and Precipices, or a wide Expanse of Waters” (Monk, 1935, p. 57). This illustration strikes an affinity with that of Gerard’s from a century later, where the sublime is associated with “the Alps, the Nile, the ocean, the wide expanse of heaven, or the immensity of space uniformly extended without limit or termination” (Costelloe, 2012, p. 65).

This is unlikely an artifact circumscribed to British worldviews, as was demonstrated in the examples before. Words related to size and height appear often as synonyms of the sublime in various languages; Arabic (**ضخم; رفيع**), Chinese (嵩高; 崇高), German (großartig; erhöhen), Persian (**رفيع; والا**), and Russian (возвышенный, грандиозный).

Size and Height, a Psychological Perspective

In psychological research, size has often been linked with liking, i.e. “bigger is better”, which may explain why people enjoy large screens (Lombard, 1995). Both adults and three-year-old participants preferred abstract objects that were larger in forced-choice tasks (Silvera, Josephs, & Giesler, 2002). The preference for size can also be observed in natural settings. Frynta and colleagues (2010, 2013), for example, reported that zoos prefer to keep large animals because they are liked by both adults and children (Ward, Mosberger, Kistler, & Fischer, 1998). As Silvera et al. (2002) noted, there may be a “fundamental, reliable association between physical size and human preference” (p. 191).

A number of recent psychological theories postulated size as an integral component of awe and similar sublime-like psychological states[2] (Keltner & Haidt, 2003; Konečni, 2005, 2011). Joye and Verpooten (2013) argued that the impressiveness of large religious buildings can largely be attributed to the sensory experience of largeness. Here, size is an aesthetic experience in and of itself,

[1] The etymology of sublimity itself implies physical elevation (Cohn & Miles, 1977).

[2] Since Keltner and Haidt’s (2003) paper on awe cites Burke’s *A Philosophical Enquiry into the Origins of our Ideas of the Sublime and Beautiful*, awe can be seen as an inquiry into sublimity. Besides, Burke often uses the work awe in describing sublimity in his treatise.

separate from indirect reverence inferred by perceived costliness. In support of these viewpoints, empirical works have reported links between awe and perceived vastness of an object or the perception of "the presence of something greater than myself" (Piff et al., 2015; Gordon et al., 2016; Ishizu & Zeki, 2014; Seidel & Prinz, 2017; Shiota, Keltner, & Mossman, 2007).

There are also various implications on the importance of height concerning sublimity-like states. Konečni (2011), for example, argued that the Great Pyramid of Giza and the Giant Buddha in Leshan are real-life examples of sublimity-inducing objects due to their prominent height. Likewise, Joye and Verpooten (2013) wrote that height is a "historically constant feature" of monumental buildings because height represents power.

While no empirical work has directly tested the role of stimulus height on the elicitation of sublimity, a small number of works have provided important insights. Meier, Hauser, Robinson, Friesen, and Schjeldahl (2007), for example, reported the association between high (vs. low) vertical positioning and the perceived divinity of stimuli. Specifically, participants attributed belief-in-God to stranger photographs located in a high vertical position. Importantly, these effects were maintained after controlling for perceived power and likability (see Giessner and Schubert [2007] for a similar approach to linking verticality and perception of interpersonal power). Similarly, Joye and Dewitte (2016) observed that high-level buildings elicit greater feelings of awe than low-level buildings. These studies point to the direction of embodied cognition, where metaphors as abstract concepts are associated with perceptual representation (Lackoff & Johnson, 1999). In the present case, the metaphor "to look up to" may hold an effect (uniquely associated with the metaphor, i.e. to admire) when people are made to physically look up to an object or to look at an object that is above something else.

To present knowledge, only a study by Seidel and Prinz (2017) has tested sublimity's relationship with both size and height via experimental manipulations. Their results demonstrated that physically enlarged and heightened artworks are judged as more wonderful. Conversely, the more wonderful an artwork is seen, the greater its estimated presentation dimension and height. The implications are that an object's presentation size and height are core to sublimity experiences.

Size Perception and Visual Angle

What are the mechanisms of the aforementioned size effect? Size perception relies on three factors: 1) actual size of an object, 2) retinal image size, or the angle an object subtends at the eye, or visual angle, and 3) viewing distance. With any of the two variables, the other is automatically determined[3] (Holway & Boring, 1941; Thouless, 1931).

The importance of the three size factors in understanding size perception notwithstanding, existing psychological studies manipulating object size perception have tended to overlook systematic control of these three factors. Studies commonly attribute size effects to changes to stimulus size, even though the effect could have also been a result of changes to visual angle – when the viewing distance is kept constant, visual angle and absolute size are confounded with each other. Thus, Seidel and Prinz's (2017) conclusion that the perception of wonder in an artwork increases with the artwork's size is inconclusive since the effect may have been an effect of stimulus visual angle as well as actual stimulus size.

Similar imprecisions have also occurred in philosophy. Burke, like Seidel and Prinz (2017), did not venture into the possibility of large objects being viewed from various distances. Accordingly, the mechanism of the size-induced account of sublimity is blurred in Burke's arguments. This potential confound is reflected in Payne Knight's comments on Burke, that "one's pen a foot away makes a greater impression on the retina than Salisbury steeple at a mile, and the sheet of paper on which one writes would be more sublime than the Peak of Teneriffe" (Hipple, 1957, p. 92). Research into visual size, actual size, and viewing distance, thus, may provide crucial clues of the workings of the previously observed size effects.

Color

While less prominent than size and height, color may be another important feature in discussions of sublimity and beauty. The philosophical literature, particularly Burke's *A Philosophical Enquiry* (1759/2008), suggests three features, namely colorfulness, brightness, and contrast.

[3] $V = 2\tan^{-1}\frac{S}{2D}$. V is the visual angle; S, the actual size of the object; and D, the distance between viewer and object.

According to Burke, an object is beautiful "if the colors be strong and vivid, they are always diversified, and the object is never of one strong color" (p. 116). In doing so, Burke views colorfulness as a unique predictor of beauty but not of sublimity. Burke's explanation is a physiological one, where "rays frequently vary their nature, now to blue, now to red, and so on" (p. 136), result in feelings of calmness and beauty. On the other hand, sublimity is evoked through blackness, blackness being something "[that] cannot be considered as a color" (p. 145).

Brightness and contrast may also be associated with sublimity and beauty. Burke (1759/2008) wrote, "darkness is more productive of sublime ideas than light" (p. 79-80), given that darkness creates obscurity and terror. Anything bright, on the other hand, may be source of the beautiful, which roots from elicitations of "mere positive pleasure" (Burke, 1759/2008, p. 158).

A close reading of *A Philosophical Enquiry* (1759/2008) reveals Burke's consideration of contrast as an equally important source of the sublime. In explaining how one can conjure sublimity in architecture, for example, he suggests that one "ought to pass from the greatest light to as much darkness" (p. 81). In Burke's citation of Milton's depiction of Deity in *Paradise Lost*, "Dark with excessive light thy skirts appear" (p. 80), the conjured imagery indicates heightened contrast. On the other hand, given Burke's conception of beauty operating on properties of gradual variation, lack of abruptness, and softness, it can be postulated that strong contrast may oppose beauty.

Color, a Psychological Perspective

Colorfulness has occasionally been associated with subjective beauty. For instance, when rating abstract patterns, subjective beauty was associated with perceived colorfulness and manipulated colorfulness (Jacobs et al., 2016). Relatedly, colored objects, compared to their grayscale versions, were seen as more likable (Labreque & Milne, 2012) and more aesthetic (Massaro et al., 2012).

Yet there are contrasting findings that argue that color may not have aesthetic functions. Forsythe, Nadal, Sheehy, Cela-Conde, and Sawey (2011) reported no difference of subjective beauty between colored and grayscale versions of photographs and artworks. Similarly, Lyssenko, Redies, and Hayn-Leichsenring (2016) reported that subjective beauty has no relation to color saturation (i.e. colorfulness) in abstract paintings. A recent work by Reymond, Pelowski, Opwis, Takala, and Mekler (2020) reported a positive effect of color saturation on liking ratings of digital paintings, especially

among laypeople. Color saturation manipulation did not affect overall ratings of "being moved", although this was moderated by one's art viewing expertise (i.e. lay viewers were moved than experts). Color also did not affect the cropping of photographs (Mathers, 2013; McManus, Zhou, l'Anson, et al., 2011).

Empirical works have also reported brightness's association with preference (Che, Sun, Gallardo, & Nadal, 2018). There have been reported that high contrasted images are aesthetically pleasing (van Dongen & Zijlmans, 2017) and pretty (Reber, Winkielman, and Schwarz, 1998). Given Burke's (1759/2008) association of pleasure and prettiness with beauty, these findings point to the direction that brightness may be associated with beauty judgments.

Sublimity and Beauty, a true Dichotomy?

The literature presented thus far portrayed sublimity and beauty as opposing experiences, each operating on their own visual triggers. Yet, this premise is not free from challenges. In philosophy, Addison – to an extent, the forefather of British debates on the sublime – outlined sublimity and beauty as just two types of multiple aesthetic experiences, meaning that sublimity and beauty may be separate experiences but not necessarily opposite experiences. On the other hand, the likes of Baillie and Sir Joshua Reynolds saw many commonalities in sublimity and beauty, with the latter seeing sublimity as a subcategory of beauty (Hipple, 1957). This last viewpoint is echoed by Konečni (2011), who viewed sublime experiences as a kind of a beauty experience. Lastly, in contrast to the reviewed literature, there have been theoretical viewpoints on how sublimity can be elicited through small objects or by peering down onto objects from above (Burke, 1759/2008; see also Schopenhauer) and how physically great objects can elicit beauty (Hipple, 1957)

The notion that sublimity and beauty may be related to each other, in fact, is also reflected in the earlier-discussed empirical papers. While Ishizu and Zeki (2014) successfully demonstrated that sublimity and beauty elicit different areas of the brain (note, however, that beauty activations were based on painting rating and sublimity activations were based on photograph rating), sublimity and beauty ratings were positively correlated nevertheless, with substantial effect size. Similarly, despite Hur et al. (2020) replicating, using two independent samples, the fact that sublimity and beauty have different emotional profiles, sublimity and beauty were likewise positively correlated in both samples.

Furthermore, in a paper by Pelowski and colleagues (2021), when people were asked about personal sublime experiences, a number of cases involved small objects (e.g. fireflies); importantly, regardless of their physical elicitors, sublime experiences were commonly associated with experiences of beauty and pleasure, as well as negative emotions. It is also worth noting that empirical works have often cited blue as the color that is most associated with sublimity (Pelowski et al., 2021) as well as beauty (Komar & Malamid, 1999), in contrast to Burke's focus on dark colors in eliciting sublimity.

As such, the sublimity-beauty dichotomy presents ambiguities, and the possibility that sublimity and beauty are different but also similar cannot be precluded. Despite such subtleties, the literature review mainly covered arguments surrounding Burke's *A Philosophical Inquiry*, given its widespread, popular influence on subsequent frameworks of major sublime theories (e.g. Kant), and given its uniquely empirical and experiential interpretation of sublimity and beauty.

Overview of Current Research

To sum up, despite the psychological literature that often overlooks the relationship between sublimity and beauty, there are reasons to believe that sublimity and beauty may operate on selective visual mechanisms. In other words, should responses (in this case, sublimity and beauty) selectively change due to specific visual manipulation, this may imply that the responses may operate on separate underlying mechanisms. In this context, the reviewed literature discussed how sublimity has been linked with largeness, height, darkness, and high contrast, while beauty has been associated with smallness, color, and brightness. If the manipulation of specific visual features impacts sublimity positively but not beauty (or beauty negatively) and other visual features impact beauty positively but not sublimity (or sublimity negatively), this may provide evidence for a dissociation, meaning that sublimity and beauty may be based on differing psychological mechanisms (see Dunn & Kirsner, 2003, for an overview of dissociations in psychology). The present paper is the first paper to systematically explore this idea.

The paper is divided into three empirical studies. These three studies are preceded by a pilot studies section and followed by an aggregated data analysis. The three numbered studies followed similar experimental settings, where participants rate photographs projected onto a wall.

Specifically, in the pilot studies, participants rated a large set of photographs in preparation for stimuli selection for subsequent studies. In Study 1, participants rated the sublimity and beauty of photographs that are manipulated in presentation size and height. In Study 2, participants rated photographs based on their manipulated colorfulness and size. In Study 3, size and color variables were further explored, via manipulation of visual angles, viewing distance, brightness, and contrast. The aggregated data analysis combined data from all three studies to further generalize the previous size effects.

The work also proposes methodological advancements. Empirical works on the sublime often draw conclusions using a small number of stimuli (e.g. Seidel and Prinz, 2017). On these occasions, generalizability and statistical power are compromised despite sufficient participant numbers (Judd, Westfall, & Kenny, 2017; Westfall, Kenny, & Judd, 2014). After all, there is little guarantee that an effect found using a single Picasso painting can be replicated across different paintings, let alone other paintings by Picasso. Overall, 233 photographs and 245 participants are considered in total throughout the paper. Analytically, linear mixed models were used, which enables maximal consideration of random effects structures pertinent to various variations across stimuli and participants, improving upon generalizability of fixed effects (Judd et al., 2017).

Pilot studies: Stimuli Generation and the *Aesthetic Hexagon*

A set of pilot studies (approved by University College London's Research Ethics Committee) were run as preliminary explorations into rating behaviors of sublimity and beauty. Crucially, the task served the purpose of stimuli selection for the subsequent empirical studies. Across 96 participants, a total of 113 unique photographs were rated for their elicited degrees of both sublimity and beauty. Photographic content, derived from the National Geographic website and from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997), was diversified as much as possible. The content diversification meant that photographic content was repeated as little as possible throughout the stimuli set. Photographic contents were only repeated in the event of notable, stylistic differences, e.g. stormy sea vs. calm sea (see Supplemental Material for further details about the pilot studies).

Based on aggregated ratings across participants, there was a positive correlation between sublimity and beauty ratings ($r = 0.78$, $p < .001$). Yet further exploration of the data indicated that the images that were on the outer edges of the sublimity-beauty ratings were consistent across participants. Based on these rating behaviors, a conceptual map of sublimity-beauty ratings was introduced, namely the *Aesthetic Hexagon*.

The *Aesthetic Hexagon* represents six conceptual corners from a two-dimensional sublimity-beauty rating space, not unlike the RIASEC model in vocational psychology (Holland, 1973). Figure 1 presents a visualization of the map and lists some of the typical photograph contents per category.

In the *Aesthetic Hexagon*, the Tender/Adorable and Peaceful/Elegant categories represent two variants of stimuli considered to be more beautiful than sublime. The Powerful/Imposing, and Dreadful/Fearful categories, on the other hand, represent two variants of stimuli rated as more sublime than beautiful. The Boring/Disgusting and Marvellous/Astonishing categories represent photographs that are considered not sublime and not beautiful, and very sublime and very beautiful, respectively.

The adjective labels used to name the hexagon corners were based on a previously run word association task. In this task, 67 participants rated 112 words/phrases on a scale of 1-9 for each word/phrase's perceived degree of association with sublimity and beauty (the words/phrases were mainly selected from commonly appearing aesthetic adjectives from aesthetics texts personally selected by the primary author, but also included adjectives to denote general aesthetic and emotional states[4]). Similar to the photographic *Aesthetic Hexagon*, albeit to a much lesser extent, the words/phrases could be interpreted in a hexagonal shape. By transposing the word-based hexagon over the aforementioned picture-rating hexagon, six pairs of words were selected from the former hexagon thought to best fit and represent each corner of the latter *Aesthetic Hexagon* (see Supplemental Material for the full list of words/phrases).

[4] The aesthetic and emotional adjectives were selected based on an unpublished adjective list personally shared by Helmut Leder and Matthew Pelowski. Parts of the adjective list have since been published in various original research articles (e.g. Pelowski, Markey, Forster, Gerger, & Leder, 2017; Pelowski, Specker, Gerger, Leder, & Weingarden, 2018; Pelowski, Leder, et al., 2018). We are grateful to these authors for having shared these adjectives us.

In the empirical studies that follow, each participant is exposed to equal numbers of stimuli from each hexagon category. Furthermore, each participant will be rating subsets of stimuli (for each hexagon category) from a larger pool of stimuli. From here, two important methodological implications followed. Firstly, the diversification of stimulus for each participant means that any reported visual manipulation effect across stimuli cannot be resulting from a specific group of stimuli within the sublimity-beauty space. Therefore, there are improved generalizability for any manipulation (fixed) effects. Secondly, each participant will be exposed to slightly different sets of stimuli, even though every participant will be rating stimuli that cover a wide range within the sublimity-beauty space. This indicates that the previous point's generalizability enhancement of fixed effects can be achieved across a large number of participants and stimuli, especially without burdening an individual participant with large numbers of rating trials.

It should be emphasized that the six categories were not considered as discrete categories in the sense of categorical perception in linguistics. Rather, it was assumed that aesthetic experiences are continuously graded and that the hexagon is foremost a rough illustration of the relative positions of photograph contents within the sublimity-beauty rating relationship. During stimulus selection, the *Aesthetic Hexagon* was used as a heuristic to enable a relatively – not absolutely – diverse stimulus content selection. For these reasons, the hexagon categories were not used as categorical and/or fixed predictors throughout all analyses.

Study 1: The Roles of Stimulus Size and Height on Sublimity and Beauty Judgments

Methodology

Participants. 32 participants (16 female, M_{age} = 26.53, SD_{age} = 18.39) from University College London were recruited in return for course credits. All participants provided written consent prior to the start of the study.

Materials. Each participant looked at a subset of a pool of 60 photographs. The pool was divided into two groups, i.e. Set A and Set B, each consisting of 30 photographs. Photographs from Set A consisted of pre-rated stimuli from the pilot studies, and – as mentioned in the previous pilot studies section – consisted of a mix of IAPS and National Geographic photographs. For Set B

photographs, 22 photographs were taken from Study 1 of the Hur et al. (2020) paper, but eight were newly selected for the present study.

Unlike Set A photographs, Set B photographs were exclusively nature-based in content, without humans or artifacts. This was done to reflect the recurring theme of nature among various theories of the sublime (e.g. Burke, 1759/2008). For each of Set A and Set B, each corner of the *Aesthetic Hexagon* was represented in equal numbers of stimuli. Where ratings were not available for the newly selected Set B photographs, these newly selected eight photographs were approximated to represent one of the *Aesthetic Hexagon* by content.

Each experimental session took place in a dark room at University College London and was run via MATLAB 2016b (MathWorks, Inc., Natick, MA, USA) of a laptop. The laptop was connected to a 200W Epson EBX03 projector, and photographs were projected onto a blank wall. Participants used a computer mouse to input the photograph ratings.

Design. There were two types of photographic manipulation conditions, namely presentation size and presentation height. There were two experimental blocks in total, and each manipulation was allocated to one of the blocks. For the size manipulation (i.e. size block), photographs in the 'small' condition were presented 75 × 100 cm in dimension, whereas photographs in the 'large' condition were presented 150 × 200 cm in dimension. Both size conditions had photographs presented at eye-level (i.e. center), the photograph's center approximately 135cm raised from the floor.

For the height manipulation (i.e. height block), photographs in the 'center' condition were presented 135cm raised from the floor, whereas photographs in the 'high' condition were presented 195cm raised from the floor. Both height conditions had photographs presented in the 'small' size. Participants sat 390cm away from each stimulus.

Therefore, each photograph was presented as one of three presentation conditions, namely 'small-central', 'large-central', and 'small-high.' Note that due to the absence of a 'large-high' condition, the manipulations are not crossed across stimuli. The 'large-high' condition was left out, since the 'large' version of photographs already covered most of the presentation screen.

All participants rated both the size and height blocks (and in this order). The experimental manipulations were thus done within participants, such that each participant was subject to both size

and height manipulations. Stimulus-wise, a stimulus was presented as part of either the size or height manipulation condition. Within a block, 18 unique photographs (randomly selected from the image pool, but representing each *Aesthetic Hexagon* category in equal number) appeared twice, once for each condition. Half of the participants had Set A appear in the size block and Set B appear in the height block; the other half had the two image sets switched between the blocks.

In the first half of each block, no photograph contents were repeated, making the first 18 trials of each block a between-participants design regarding stimuli. This also meant that participants did not rate the same photograph content, albeit in a different manipulation condition until they saw all photographic content within each block – this would minimize the potential habituation effects of repeated exposure of identical photographic content. For each block, the stimuli content was presented in randomized order and in randomized manipulation conditions.

Procedure. Before the start of each study, participants were given a set of definitions of sublimity and beauty – the two dependent variables of the study. The set of definitions was identical to the one used in Hur et al. (2020).[5]

Each participant first went through a practice block of rating six photographs (not included in the photograph pool) for sublimity and beauty. That was followed by the size block, then the height block. For each trial, a fixation screen in the form of the response screen appeared for 0.5 seconds, before the stimulus appeared. Participants observed the stimulus for as long as they wished, before clicking the mouse. Participants could then rate the stimulus's elicited degree of sublimity and beauty at their own pace. This was followed by a 'next trial' screen, where participants could click their mouse whenever they were ready for the next trial. Figure 2 presents a visualization of a single trial procedure.

When rating each photograph, participants used a variant of the *Evaluative Space Grid* (Larsen, Norris, McGraw, Hawkley, & Cacioppo, 2009), as has been used in Hur et al. (2020). Here, Likert-like scales of sublimity and beauty were simultaneously presented in the x (measuring

[5] Sublimity was characterized as: "The degree to which one feels in an object a sense of power, loftiness, shock and/or the grandiose. Synonyms: awe-inspiring, grand, great, immense, vast and/or imposing." Beauty was characterized as: "The degree to which one feels in an object a sense of pleasure, elegance and/or joy. Synonyms: attractive, pretty, or loving."

sublimity) and y (measuring beauty) axes of a gray square. By clicking once in an area within the square using a computer mouse, each participant thus gave ratings of the two scales simultaneously.

The research was approved by University College London's Research Ethics Committee.

Results

Manipulation checks. To assess the validity of the ratings themselves, a set of manipulations checks were run. There were acceptable levels of test-retest reliability, between-studies reliability, and between-participants agreement for both sublimity and beauty ratings (see Supplemental Material).

Presentation size and height effects on sublimity and beauty.

Analytical approach. In determining the size and height presentation effects, the size and height effects were analyzed separately. For each analysis, ratings were predicted by manipulation type (i.e. small vs. large or center vs. top) and judgment type (i.e. sublimity vs. beauty), and their interaction. The interaction term enables assessing the degree to which the presentation manipulation affects sublimity and beauty differentially. A second set of models were run, this time predicting one response type (e.g. sublimity) whilst keeping the other as a covariate (e.g. beauty). This second analysis represented the fixed effect accounting for any correlation between sublimity and beauty.

Inferential statistics were calculated using linear mixed models, where random effects structures are expanded to incorporate variances deriving across both participants and stimuli simultaneously (traditional statistical models usually treat only participants as the random factor; Judd et al., 2017; Westfall et al., 2014). Since linear mixed models permit generalizations of fixed effects that are beyond sampled participants and stimuli, and since empirical aesthetics research often attempt to generalize fixed effects using complex stimuli and higher-level responses that likely accompany high degrees of variances among participants and stimuli, linear mixed models are becoming increasingly useful in aesthetics (Hur et al., 2020; Vartanian et al., 2019). In other words, the use of linear mixed models enables fixed effects that are better controlled for Type 1 error and are more generalizable in empirical aesthetics research.

All linear mixed model analyses in the paper were computed using the *lmer()* function of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in *R* version 3.4.1 (R Development Core Team, 2008). Specifically, each linear mixed model was constructed with maximal random effects

structure (with random slopes and intercepts), in accordance with the guidelines by Barr, Levy, Scheepers, and Tily (2013) and Judd et al. (2017). This method reflected the way linear mixed models are implemented in empirical aesthetics works (Hur et al., 2020; Vartanian et al., 2019). In the event of convergence errors, higher-order interaction terms in random effects were dropped, as recommended by Judd et al. (2017). A sample code with further explanation can be found in the Supplemental Material.

p-values at 95% confidence intervals for all fixed effects (Type II) and pairwise comparisons were obtained using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). Pairwise comparisons were based on least-square means. Unless specified otherwise, estimations of *t* and *F* statistics were based on the Satterthwaite approximation for degrees of freedom. Unless specified otherwise, mean values represent least-square means (i.e. estimated marginal means) derived from the *lmerTest* package.

Size effect. For the size manipulation block, there was a main effect of size (M_{small} = 0.43; M_{large} = 0.53; F[1, 49.74] = 53.45, $p < .001$), with large photographs rated higher in both sublimity and beauty than small photographs. There was no main effect of judgment type ($M_{sublimity}$ = 0.50; M_{beauty} = 0.46; F[1, 68.62] = 0.73, $p > .05$), demonstrating that there were, in general, no differences between sublimity and beauty ratings over both size conditions.

There was a significant interaction between manipulation type and judgment type (F[1, 81.53] = 19.51, $p < .001$). The interaction meant that although ratings of sublimity and beauty similar when presented small ($M_{sublimity}$ = 0.44; M_{beauty} = 0.43; t[69.90] = 0.27, $p > .05$), relatively more sublimity was evoked than beauty when images were presented large ($M_{sublimity}$ = 0.57; M_{beauty} = 0.50; t[65.10] = 1.80, $p = .08$). Increase of stimulus size increased both sublimity (t[50.60] = 7.65, $p < .001$) and beauty (t[47.10] = 5.18, $p < .001$). Descriptive and inferential statistics are provided in Table 1. The outcomes are visualized in Figure 3.

Sublimity ratings were predicted by the size manipulation, with beauty ratings set as a covariate (M_{small} = 0.45; M_{large} = 0.55; F[1, 48.60] = 46.25, $p < .001$). Conversely, when beauty ratings were predicted by the size manipulation with sublimity ratings set as a covariate, beauty was also predicted by size (M_{small} = 0.44; M_{large} = 0.48; F[1, 46.60] = 13.45, $p = .001$).

Height effect. For the height manipulation block, there was a main effect of height *(M_{top} = 0.48; M_{center} = 0.43; F*[1, 38.50] = 20.57, $p < .001$), meaning that photographs positioned at the top, compared to those positioned centrally, were rated higher in both sublimity and beauty. There was no main effect of judgment type *($M_{sublimity}$ = 0.46; M_{beauty} = 0.45; F*[1, 68.52] = 0.14, $p > .05$), meaning that there were no differences between sublimity and beauty ratings over both height conditions. Likewise, the lack of a significant interaction between manipulation and judgment types (F[1, 38.08] = 0.42, $p > .05$) meant that the height effect did not affect sublimity and beauty differentially. Descriptive and inferential statistics are provided in Table 1.

Even after controlling for the other rating variable, the height effect predicted sublimity (M_{top} = 0.49; M_{center} = 0.44; F[1, 47.42] = 8.89, p = .005), and beauty (M_{top} = 0.46; M_{center} = 0.43; F[1, 32.09] = 5.72, p = .02).

Discussion

The present study explored the effects of presentation size and height on sublimity and beauty elicitation through photographs. Where an increase of presentation size or height of photographs resulted in higher elicitations of sublimity and beauty, size affected sublimity more than beauty. Height, on the other hand, affected sublimity and beauty to similar degrees. It was further determined that the ratings of sublimity and beauty while being moderately and positively related, had good test-retest and between-cohort reliabilities, and between-participants agreement. Study 2 was run to further generalize the size effect from Study 1 using some newly selected stimuli and participants, and – given the literature on the role of colorfulness on aesthetic judgments – to test the additional role of colorfulness on sublimity and beauty.

Study 2: The Roles of Stimulus Size and Color on Sublimity and Beauty Judgments

Methodology

Participants. As in Study 1, Study 2's data collection was done in return for course credit at UCL. 39 participants took part (24 female, M_{age} = 20.08, SD_{age} = 1.01). All participants provided written consent prior to the start of the study.

Materials. Each participant looked at a subset of a pool of 60 photographs – these photographs were previously rated either in Study 1 of the present paper or in Study 2 of Hur et al. (2020). As with Study 1, the photographs were chosen to approximately represent the *Aesthetic Hexagon* in equal stimulus numbers, with ten unique photographs for each of the six categories.

Design. The design of the study was a 2 (size: small vs. large) × 2 (color: color vs. monochrome) within-participants design, as each participant observed all four of the stimulus manipulations. Therefore, the present design enabled the estimation of an interaction between the two main manipulations.

The size condition manipulation was identical to that of the previous study. The color condition manipulation of converting colored photographs to a grayscale monochrome was done via MATLAB 2016b's (MathWorks, Inc., Natick, MA, USA) *rgb2gray()* function.

Regarding the main rating section of the study, 30 unique photographs from the entire stimulus pool were semi-randomly selected for each participant. Unlike Study 1, the photographs were not allocated to two different sets, allowing for purer randomization in stimuli selection per participant. The selected 30 photographs for each participant were composed of five photographs from each corner of the *Aesthetic Hexagon*. Of the five photographs from each hexagon corner, either two or three (randomly decided) photographs were presented in color, leaving the other three or two in monochrome.

For each participant, all 30 photographs appeared once in large and once in small, but in the same color condition. For example, a photograph chosen for a participant in color would appear once as colored-large and once as colored-small throughout a session, but never in monochrome. This design prevented the participants from repeatedly seeing all four manipulation conditions of a photograph content.

Similar to Study 1's design, until halfway point of a session, no stimuli contents were repeated. Until then, the appearance order of photograph content and size manipulation, and color manipulation were randomized. Therefore, up until the mid-point of the entire image rating task, the design adhered to a between-participants design in terms of the stimulus content. There was a small break halfway through the 60-trialed main section of the study.

Procedure. Unlike in Study 1, participants viewed each stimulus for six seconds instead of viewing a stimulus for as long as they wished. By better controlling stimuli exposure time across participants and stimuli, potential psychological consequences deriving from varying stimuli exposure time were controlled for.

The research was approved by University College London's Research Ethics Committee.

Results

Manipulation checks. As in Study 1, there were good levels of test-retest reliability, between-studies reliability, and between-participants agreement (see Supplemental Material). The positive between-study reliability implies that image viewing time (a key methodological difference between Studies 1 and 2) did not seem to affect one's overall sublimity and beauty ratings in the present research context.

Presentation size and color effects on sublimity and beauty. Ratings were predicted by the main effects of size manipulation, color manipulation, judgment type, and all possible interactions between these variables. There were main effects of manipulations of size (M_{small} = 0.50; M_{large} = 0.53; F[1, 85.70] = 19.24, p < .001) and color (M_{color} = 0.54; $M_{monochrome}$ = 0.50; F[1, 41.80] = 14.33, p < .001). That is, both sublimity and beauty ratings increased as a photograph was generally presented as large as opposed to small and presented in color as opposed to monochrome.

There was a significant interaction between color and judgment type (F[1, 3493.80] = 7.53, p = .006), meaning that the effects of color manipulation had selective effects on sublimity and beauty ratings. Although photographs presented in monochrome lead to similar sublimity and beauty experiences ($M_{sublimity}$ = 0.50; M_{beauty} = 0.49; t[63.90] = 0.40, p > .05), photographs presented in color are more beautiful than they are sublime ($M_{sublimity}$ = 0.53; M_{beauty} = 0.56; t[68.50] = 2.30, p = .03). Similarly, where colored photographs elicit substantial more beauty than monochrome photographs (t[68.50] = 4.63, p < .001), this color effect is relatively muted for elicitations of sublimity (t[68.50] = 2.06, p = .04). No other main or interaction effects reached statistical significance at p < .05. Descriptive and inferential statistics are provided in Table 2. The outcomes are visualized in Figure 3 (see Supplemental Material for visualization of full model).

While Study 1's findings, namely the interaction between presentation size and judgment type was not replicated (F[1, 3458.00] = 1.22, $p > .05$), the interaction effect did not reach significance even when a model was re-fitted only using presentation size, judgment type, and their interaction as predictors, (F[1, 3371.8] = 2.00, $p > .05$).

As with Study 1, the sublimity and beauty ratings were predicted by each fixed effect one at a time while the other rating variable was held as a covariate. After controlling for beauty, sublimity was predicted by size (M_{small} = 0.49; M_{large} = 0.53; F[1, 40.68] = 20.52, $p < .001$), but not by color (M_{color} = 0.52; $M_{monochrome}$ = 0.50; F[1, 114.53] = 2.07, $p > .05$). After controlling for sublimity, beauty was predicted by size (M_{small} = 0.52; M_{large} = 0.54; F[1, 37.09] = 6.78, p = .01), and color (M_{color} = 0.56; $M_{monochrome}$ = 0.49; F[1, 46.08] = 20.60, $p < .001$). No other effects were significant. These outcomes emphasizes that color seems to predict beauty ratings, but not sublimity ratings (see Supplemental Material for full output table).

Discussion

Study 2 explored the roles of presentation color and size on their effects on reported sublimity and beauty. While color and size both increased ratings of sublimity and beauty, color affected beauty more than sublimity. In both Studies 1 and 2, size was associated with positive aesthetic evaluations for both sublimity in beauty. Although Study 1 reported that size affects sublimity more than beauty, this was not replicated in Study 2.

What are the mechanisms of the size effect? In Study 3, the size effect was further explored via breaking down stimulus size into visual angle, viewing distance, and actual size. Furthermore, two further color variables, namely brightness and contrast were additionally considered.

Study 3: The Roles of Visual Angle, Viewing Distance, Brightness, and Contrast, on Sublimity and Beauty Judgments

judgment

Methodology

Participants. 39 participants (28 female, M_{age} = 22.64, SD_{age} = 10.24) from University College London were recruited in return for course credit. All participants provided written consent prior to the start of the study.

Materials. Each participant looked at a subset of a pool of 102 photographs in landscape orientation. 12 of these photographs were taken from Study 2, of which two photographs were from each *Aesthetic Hexagon* corner (see the introduction for a description of the *Aesthetic Hexagon*). Of the 90 newly selected photographs, 30 of them were chosen from the IAPS database (Lang et al., 1997), while the rest were newly selected from online sources (YG & YH). As with previous studies, these newly selected photographs approximated to represent the six corners of the *Aesthetic Hexagon* in equal numbers, and the photograph contents were diversified as much as possible.

Design. For each participant, a random selection of 12 photographs was taken from each *Aesthetic Hexagon* category of the photograph pool, amounting to a total of 72 unique photographs per experiment session. Unlike previous studies, a specific photograph content appeared only once per participant. Like the past studies, all participants went through all manipulation conditions. The design of manipulation conditions is described below. In this new design where stimuli content was not nested within specific manipulation conditions, fixed effects of the manipulations were generalizable across image contents.

Visual angle and viewing distance. For the size manipulation, the 72 photographs were semi-randomly assigned in equal numbers to a 2 (viewing distance; 57cm vs. 390cm) × 2 (visual angle; 10.98˚ × 14.64˚ vs. 21.77˚ × 38.70˚) within-participants factorial design. The allocation was semi-random because each of the four size conditions was planned to contain equal numbers of photographs randomly selected within each *Aesthetic Hexagon* category. The design, thus, resulted in each participant viewing four actual size conditions: 150cm × 200cm, 75cm × 100 cm, 21.92cm × 29.23cm, and 10.96cm × 14.61cm. Note that the first two conditions are replicated in Studies 1 and 2.

The close distance (57cm) conditions were presented on a computer screen, whereas the long-distance (390cm) conditions, as in the two previous studies, used a projector (200W Epson EBX03) to project images onto a wall. A photometer (Tacklife LM01) was used to calibrate the amount of light

emitting from the computer screen and wall. Two white rectangles in the size of images of the small and large visual angle conditions were put onto the screen and the wall, respectively. The screen's settings were adjusted so that the small rectangles on the screen and wall both emitted 4.1 lux from both viewing distances. The large rectangles on the screen and wall both emitted 17.5 lux from both viewing distances.

Brightness and contrast. All 72 images per session were entirely randomly subjected to one of four brightness manipulations, namely 1) high brightness, 2) low brightness, 3) high contrast, 4) low contrast, in equal numbers. bmp_contrast.m in MATLAB 2016b (MathWorks, Inc., Natick, MA, USA) was used to manipulate the brightness and contrast of the images. The global brightness of an image was calculated using the logarithmic average of luminance (Y) from the YUV system, as suggested by Li and Chen (2009). The global contrast of an image was calculated via the root mean square (RMS) of Y of that particular image (Peli, 1990). Images in the two brightness conditions (i.e. low vs. high brightness) were matched in contrast. Likewise, images in the two contrast conditions (i.e. low vs. high contrast) were matched in brightness (the average brightness and contrast levels across the four conditions are available in the Supplemental Material). Please note that the brightness and contrast manipulations were not constructed as a full factorial design, since an image manipulated both in brightness and contrast substantially compromised the naturalness and realism of the stimuli. Given the aim to retain the naturalness of all stimuli, the stimuli were designated either to one of the four brightness or contrast conditions.

Procedure. The general setup and procedure of the experiment were identical to those of the two past studies, apart from the following changes. There were two main blocks, one using the computer monitor (short viewing distance condition) and one using the wall (long viewing distance condition). Half of the participants attended the short viewing distance condition first, whereas the other half of the participants took part in the long viewing distance condition first. Within each block, the two visual angle conditions (small vs. large visual angle) and the four brightness conditions were randomized in appearance order. All participants were given three practice trials prior to each block.

The research was approved by University College London's Research Ethics Committee.

Results

Manipulation checks. Because each participant looked at one image per session, test-retest reliability was not available. As with the past two studies, there were acceptable levels of between-studies reliability and between-participants agreement (see Supplemental Material).

Visual angle and viewing distance effects. Ratings were predicted by visual angle (i.e. small vs. large visual angle), viewing distance (i.e. short vs. long viewing distance), judgment type (i.e. sublimity vs. beauty), and all possible interactions between the variables. There was a main effect of visual angle ($M_{smallVA}$ = 0.48; $M_{largeVA}$ = 0.50; F[1, 2613.81] = 14.97, p < .001), indicating that the larger the visual angle, the greater the general ratings of sublimity and beauty, regardless of the viewing distance condition. The descriptive and inferential statistics output are presented in Table 3.

The interaction between visual angle and judgment type (F[1,2673.53] = 5.79, p < .05) was derived from the fact that sublimity ratings were higher for photographs presented in large visual angle than small visual angle ($M_{smallVA}$ = 0.46; $M_{largeVA}$ = 0.49; t[5221.40] = 4.48, p < .001), whereas for beauty ratings, photographs presented in large visual angle were just as beautiful as photographs presented in small visual angle ($M_{smallVA}$ = 0.50; $M_{largeVA}$ = 0.51; t[5219.90] = 1.26, p > .05). When photographs were presented in small visual angle, beauty was not higher than sublimity (t[130.60] = 1.53, p > .05). Likewise, sublimity and beauty ratings did not differ when photographs were presented in large visual angle (t[130.60] = 0.60, p > .05). When a model was re-fitted to have visual angle, rating type, and their interaction predict ratings, the results were replicated (see Supplemental Material for visualization of the full model).

When sublimity was predicted by visual angle, distance, and their interaction, with beauty as a covariate, sublimity was predicted by visual angle only ($M_{smallVA}$ = 0.46; $M_{largeVA}$ = 0.49; F[1, 68.53] = 13.38, p < .001). When beauty was predicted by the three main predicting variables, with sublimity as a covariate, no predictors predicted the outcome variable, including visual angle ($M_{smallVA}$ = 0.50; $M_{largeVA}$ = 0.51; F[1, 80.59] = 0.72, p > .05).Thus, visual angle appears to affect sublimity but not beauty, and viewing distance does not seem to play a major role (see Supplemental Material for full output table).

Lastly, the role of actual image size was estimated by comparing ratings of the same visual angle conditions in different distances. When this was done for all possible comparisons, no

comparison was significant at $p < .05$, indicating that actual size did not play a role in determining ratings of sublimity and beauty.

Brightness and contrast effects. In a single model, ratings were predicted by brightness/contrast manipulation (i.e. low brightness vs. high brightness vs. low contrast vs. high contrast), judgment type (i.e. sublimity vs. beauty), and the interaction between them. The brightness and contrast manipulations were put in as a single factor since all images appeared as one of the four manipulations. The only significant effect was the judgment type main effect ($M_{sublimity} = 0.48$; $M_{beauty} = 0.50$; $F[1, 5328.10] = 19.31$, $p < .001$), with sublimity being generally rated lower than beauty across conditions (See Supplemental Material for outcome tables and visualizations).

Discussion

The interaction between size and judgment type was present as has been the case for Study 1. The effect seemed to be explained by visual angle, as viewing distance played no substantial role. There was also little evidence of the importance of actual image size on ratings. This means that an image being viewed as small or large on the retina seems the most relevant determinant of any size effect on sublimity and beauty. When considering the roles of brightness and contrast on ratings, brightness and contrast played negligible roles in determining sublimity and beauty ratings.

To what degrees can the findings regarding visual angle be generalized across multiple studies? An aggregated data analysis was run combining data from Studies 1, 2, and 3. In the involvement of a larger number of participants and stimuli, a more statistically powerful and generalizable set of results was made.

Aggregated Data Analysis

Methodology

Datasets. The aggregated dataset comprised of data from Studies 1, 2, and 3. Additionally, data from a study that was run in a similar setting (but not part of the three numbered studies from the paper; see Supplemental Material for additional information about this study) was aggregated. All four studies were based on the identical experimental settings, e.g. experiment room and types of stimuli used (photographs), and involved the following two visual angle conditions of photograph

presentation: 10.98˚ × 14.64˚ (i.e. small visual angle) and 21.77˚ × 38.70˚ (i.e. large visual angle). Observations were selected where photographs were presented at eye-level, with color, and without the simultaneous presence of other modalities. This resulted in 7770 valid trials (3120 small visual angle trials), spread across 168 unique photographs and 149 participants.

Analytical method and variable selection. Ratings were predicted by stimulus visual angle (i.e. 10.98˚ × 14.64˚ vs. 21.77˚ × 38.70˚), judgment type (i.e. sublimity vs. beauty), and their interaction. The study number was added as a covariate, to account for potential differences across studies. The entire aggregated data were entered into a single linear mixed model.

Results

Overall visual angle effect. Using the aggregated dataset, there was a main effect of visual angle ($M_{smallVA}$ = 0.48; $M_{largeVA}$ = 0.52; F[1, 120.00] = 57.67, p < .001), indicating that the greater visual angle, the more sublime and beautiful a photograph was rated. The lack of a significant main effect of judgment type ($M_{sublimity}$ = 0.49; M_{beauty} = 0.51; F[1, 221.20] = 2.43, p > .05), implies that across the two different visual angle conditions, people tended to rate beauty similarly to sublimity.

There was an interaction between visual angle and judgment type (F[1, 7587.80] = 18.01, p < .001), meaning that the effect of visual angle was different between sublimity and beauty ratings. The increase of visual angle significantly increased both sublimity ($M_{smallVA}$ = 0.46; $M_{largeVA}$ = 0.52; t[206.90] = 8.74, p < .001) and beauty ($M_{smallVA}$ = 0.50; $M_{largeVA}$ = 0.53; t[206.10] = 4.52, p < .001) ratings, respectively. When photographs were presented with small visual angles, they were more beautiful than sublime, (t[239.60] = 2.30, p = .02). However, photographs presented in large visual angles were as sublime as they were beautiful (t[232.70] = 0.79, p > .05). As such, with the datasets from four studies put together, the role of visual angle is similar to the outcomes from Study 3. Descriptive and inferential statistics tables are presented in Table 4. The outcomes are visualized in Figure 4.

Sublimity and beauty were correlated in the overall data(r = 0.32, p < .001). To explore visual angle's effects on sublimity and beauty after controlling for this correlation, two more models were run. Each model predicted visual angle's effect on one response variable, while the remaining

response variable was taken in as a covariate. Study number was taken as a covariate in both models. In both models, visual angle predicted increase of both sublimity ($M_{smallVA}$ = 0.45; $M_{largeVA}$ = 0.51; F[1, 120.80] = 44.66, p < .001) and beauty ($M_{smallVA}$ = 0.50; $M_{largeVA}$ = 0.53; F[1, 109.10] = 20.97, p < .001). These results reflect Study 3's results and thereby further confirm visual angle's selective effects on sublimity and beauty ratings.

Discussion

In the aggregated data analysis, the overall visual angle effect from Study 3 was replicated. While the increase of stimulus visual angle appears to increase both sublimity and beauty, the effect seems more substantial for sublimity than for beauty.

General Discussion

In the present paper, the visual aesthetics of sublimity and beauty were explored by the manipulations of photograph size, height, color, brightness, and contrast. There was evidence that size affected sublimity more than beauty (Study 1), and that this is driven by visual angles instead of the actual size of a stimulus (Study 3 & Aggregated data analysis). Height affected both sublimity and beauty positively in similar degrees (Study 1), while color affected predominantly beauty (Study 2). Brightness and contrast neither affected sublimity nor beauty ratings (Study 3).

The findings demonstrate that sublimity and beauty, while related, are aesthetic dimensions that are selectively influenced by visual manipulations. They imply that sublimity and beauty may operate on selective perceptual systems. In doing so they, partly justify the past philosophical discussions where sublimity and beauty represent unique psychologies (e.g. Burke, 1759/2008).

Size Effect as a Visual Angle Effect

Despite more recent arguments, the traditional narrative of the sublime is that sublimity is elicited with the encounter of vast objects (e.g. Burke, 1759/2008). Replicating empirical evidence of the association between reported sublimity and perceived object size (e.g. Ishizu & Zeki, 2014), all studies in the present paper reported that the increase of stimulus size also increased sublimity ratings. The size-sublimity relationship can be explained by the fact that physically imposing objects represent traits of strength and power, characteristics that illustrate the psychologies of the sublime.

Yet the results also present subtleties. Where size appears to affect sublimity relatively more than it affected beauty, the increase of stimulus size also increased beauty ratings (Studies 1 & 2). That beauty, as well as sublimity, is positively linked with object size fits into the tendency to positively evaluate large things, e.g. "bigger is better" (Lombard, 1995).

In Study 2, size affected sublimity and beauty to similar degrees. Contextual factors may have been at play, where the presence of the color manipulation may have reduced the saliency of size manipulations. It may also be that aesthetic decisions through the manipulation of stimuli size are inherently difficult to replicate, perhaps owing to sample bias or slight changes of design (e.g. Troscianko et al., 2012). To remedy these possibilities, the present paper aggregated data from multiple studies and re-ran the analysis. With increased statistical power, that size, specifically an object's visual angle, affects sublimity and more than beauty was replicated.

Why does visual angle take precedence over actual size and viewing distance? According to Holway and Boring (1941), the visual angle becomes a reference for perceived size when the retinal image is the only information available, without the help of cues, e.g. distance, to infer an object's actual size (e.g. viewing through a small vision tunnel). Conversely, actual stimulus size became the determinant of size perception when distance cues were available (i.e. normal binocular observation). Because of the high task demand of the present study where complex stimuli were rated on aesthetic dimensions, participants may not have had the psychological capacity to estimate stimulus viewing distance. The assumption here is that mental processing requires resources that compete against each other; a resource-consuming task may attenuate the processing of other information. This means that the apparent size of stimuli was based on what was most immediately available, thus, what was evident on the retina.

Height and Colorfulness

That the increase of an object's presentation height (Study 1) increases reported sublimity confirms theoretical views of sublimity's close association with objects of great height (e.g. Konečni, 2011). This finding can be explained by height's association with power (Schubert, 2005). Power and force have been seen as emblematic of sublime experiences (Burke, 1759/2008).

Still, the degrees to which stimulus height affected sublimity and beauty ratings differentially was not discernible. Whether both sublimity and beauty share psychological mechanisms in this height effect remains to be seen. For example, height's positive effect on beauty might be linked with height's general association with positive valence (e.g. Meier & Robinson, 2004). As empirically reported by Hur et al. (2020) and Ishizu and Zeki (2014), beauty is closely related to emotional positivity. On the other hand, height's association with sublimity may be more driven by height's association with power (Schubert, 2005). Since valence and power are likely independent (Russell, 1980), this scenario would indicate that height's effects on sublimity and beauty may operate on independent psychological causes.

Future studies may address the mechanisms of the height effect, similar to what was done for the visual angle effect. In other words, where the height and eye-level of the viewer were not systematically manipulated, height perception may derive from absolute height of the object, height of the object relative to the viewer, and height of the object relative to other objects or the pictorial frame. In the present design, these possibilities were conflated.

Colorfulness's selective influence on beauty (Study 2) is partial toward Edmund Burke's (1759/2008). A uniquely emotional cause may be at play. While an association between colorfulness and emotional positivity is reported (e.g. Detenber & Winch, 2001), beauty is known to be associated with pleasure. However, quick conclusions must be taken with a pinch of salt. Lyssenko et al. (2016), for instance, did not find a link between subjective beauty and the presence of color.

Brightness and Contrast

Brightness and contrast, on the other hand, neither affected sublimity nor beauty. On the one hand, this can be attributed to errors in design. While brightness and contrast manipulations were done to not intrude on the naturalness of photographs, the brightness and contrast manipulations may have been too subtle for participants to notice. Therefore, effects from brightness and contrast levels could have benefitted from more noticeable manipulations of brightness and contrast. It may also be argued that the manipulations of brightness and contrast themselves could have been improved. For example, it is a known fact that the size of an object changes the relative amount of light emitted from the object. While attempts were made to control for this variance by calibrating the amount of light

emitted across the manipulations in Study 3 (where stimuli's size and brightness/contrast were manipulated), the design may merit a more detailed control over the brightness and contrast manipulations. Future research, for example, may address the roles of stimulus brightness and contrast on sublimity and beauty in a more controlled setting, without manipulations of stimulus size.

It could also be the case that brightness and contrast by nature make little impressions on sublimity and beauty. While Burke (1759/2008) identified that darkness is unpleasant, and therefore sublime, and that lightness, due to its pleasant nature, is beautiful, it may be Burke's error to misattribute lexical associations with physical realities. That is, while darkness and brightness as lexical symbols have emotional connotations (e.g. Valdez & Mehrabian, 1994), the darkness and brightness of objects may have undetectable psychological effects.

Limitations

A number of limitations of the present work can be suggested. It may be argued that rating higher-level stimuli (i.e. photographs) on complex measures present various unaccounted variances. This could mean, for instance, that the meanings of the measurement variables differed between participants. The meanings of the measurement variables may also have differed between stimuli; just as one may 'like' different pieces of garments for entirely different reasons, there may be cases where photographs of similar sublimity and beauty ratings may arouse different actual experiences.

Facing these criticisms, the present studies used linear mixed models to minimize information loss deriving from gross simplifications. By predicting roles of fixed effects after considering various between-participants and between-stimuli variances simultaneously, the works attempted to minimize error variances (e.g. Judd et al., 2017). Furthermore, fixed effects were generalized by the use of intricate randomization of stimulus – the effects of visual angle, for instance, are not limited to a certain set of stimuli. Lastly, between-subjects and between-studies agreements, and test-retest reliability were assessed and verified throughout all studies.

An area of contention may revolve around the fact that participants were given definitions of the sublime and the beautiful prior to their rating tasks. This method may be criticized that the dependent variables – sublimity and beauty –, despite their rich historical developments, were simplified, limiting the scope of their ecological implications. This contention was complemented by

the fact that the definition of sublimity included descriptors of size, meaning that this very characterization may have been a conflating factor in the size effects (note that this criticism, however, does not apply to the reported visual angle effects, given that there were no indications of visual angle among the characterization).

Yet the alternative, to rely on a participant's own definitions of sublimity and beauty, risked the possibility of the experimental researchers not understanding what is being measured (i.e. construct validity). Furthermore, the use of definitions as was used in a previously published work on sublimity and beauty, Hur et al., (2020), provided a sense of continuation in terms of research outcomes and their interpretations (i.e. external validity). Because the risk of violating such validities outweighed the risk of measuring something that is limited yet interpretable, the present methods were used.

It must be added that the definitions were intentionally general, with no example stimuli provided; it was, thus, entirely up to the participants – mostly unaware of philosophical theories on sublimity and beauty – how to rate the photographs. That a consistent set of results was derived using an interpretable measure is a particular strength of the present work. That said, future research must address the size effect of sublimity with minimum use of size-related terminologies when defining sublimity.

Conclusion

The present paper explored the visual aesthetics of sublimity and beauty. While sublimity and beauty seemed related in many ways given their shared visual elicitors, there was also evidence that they are influenced by different photographic manipulations. These findings, thus, suggest the possibility that sublimity and beauty may operate on separate mechanisms and thereby give a nod to the philosophical literature, but no doubt also unveil an equally compelling set of subtleties.

References

Augustin, M. D., Wagemans, J., & Carbon, C. C. (2012). All is beautiful? Generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica*, *139*(1), 187-201.

Ashfield, A., & De Bolla, P. (Eds.) (1996). *The sublime: A reader in British eighteenth-century aesthetic theory*. Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Burke, E. (2008). *A philosophical inquiry into the origins of our ideas of the sublime and beautiful*. Oxon, UK: Routledge Classics. (Original work published 1759).

Cohn, J., & Miles, T. H. (1977). The sublime: In alchemy, aesthetics and psychoanalysis. *Modern Philology*, *74*(3), 289–304.

Che, J., Sun, X., Gallardo, V., & Nadal, M. (2018). Cross-cultural empirical aesthetics. *Progress in Brain Research, 237*, 77-103.

Costelloe, T. M. (Ed.). (2012). *The sublime: From antiquity to the present*. Cambridge University Press.

Detenber, B. H., & Winch, S. P. (2001). The impact of color on emotional responses to newspaper photographs. *Visual Communication Quarterly*, *8*(3), 4-14.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations?. *Cortex*, *39*(1), 1-8.

Van Dongen, N. N., & Zijlmans, J. (2017). The science of art: the universality of the law of contrast. *American Journal of Psychology*, *130*(3), 283-294.

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2012). Stirring images: fear, not happiness or arousal, makes art more sublime. *Emotion, 12*(5), 1071-1074.

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*(1), 49-70.

Frynta, D., Lišková, S., Bültmann, S., & Burda, H. (2010). Being attractive brings advantages: the cas e of parrot species in captivity. *PloS one*, *5*(9), e12568.

Frynta, D., Šimková, O., Lišková, S., & Landová, E. (2013). Mammalian collection on Noah's ark: the effects of beauty, brain and body size. *PLoS One*, *8*(5), e63110.

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes*, *104*(1), 30-44.

Gordon, A. M., Stellar, J. E., Anderson, C. L., McNeil, G. D., Loew, D., & Keltner, D. J. (2017). The dark side of the sublime: Distinguishing a threat-based variant of awe. *Journal of Personality and Social Psychology*, *113*(2), 310–328.

Hipple, W. J. (1957). *The beautiful, the sublime, & the picturesque in eighteenth-century British aesthetic theory*. Carbondale: Southern Illinois University Press.

Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliffs, N J: Prentice-Hall.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *The American Journal of Psychology*, *54*(1), 21-37.

Hur. Y.-J. (2020). An empirical aesthetics of the sublime and beautiful (Doctoral thesis, University Co llege London, London, UK). Retrieved from https://discovery.ucl.ac.uk/id/eprint/10091352/18/Hur_10091352_thesis_redacted.pdf

Hur, Y.-J., Gerger, G., Leder, H., & McManus, I. C. (2020). Facing the sublime: Physiological correla tes of the relationship between fear and the sublime. *Psychology of Aesthetics, Creativity, and t he Arts*.

Hur, Y. J., & McManus, I. C. (2017). Representing the sublime in the VIMAP and empirical aesthetics: Reviving Edmund Burke's A Philosophical Enquiry into the Origins of Our Ideas of the Sublime and Beautiful: Comment on" Move me, astonish me... delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates" by Matthew Pelowski et al. *Physics of Life Reviews*, *21*, 135-137.

Ishizu, T., & Zeki, S. (2011). Toward a brain-based theory of beauty. *PloS one*, *6*(7), e21852.

Ishizu, T., & Zeki, S. (2014). A neurobiological enquiry into the origins of our experience of the sublime and beautiful. *Frontiers in Human Neuroscience*, *8*, 891.

Jacobs, R. H., Haak, K. V., Thumfart, S., Renken, R., Henson, B., & Cornelissen, F. W. (2016). Aesthetics by numbers: links between perceived texture qualities and computed visual texture properties. *Frontiers in Human Neuroscience*, *10*, 343.

Joye, Y., & Verpooten, J. (2013). An exploration of the functions of religious monumental architecture from a Darwinian perspective. *Review of General Psychology*, *17*(1), 53-68.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625.

Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion*, *17*, 297–314.

Komar, V., & Melamid, A. (1999). *Painting by numbers: Komar and Melamid's scientific guide to art* . Univ of California Press.

Konečni, V. J. (2005). The aesthetic trinity: Awe, being moved, thrills. *Bulletin of Psychology and the Arts*, *5*(2), 27-44.

Konečni, V. J. (2011). Aesthetic trinity theory and the sublime. *Philosophy Today*, *55*(1), 64-73.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1548–7660.

Labreque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science*, *40*, 711–727.

Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenges to Western thought. New York: Basic Books.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39-58.

Larsen, J. T., Norris, C. J., McGraw, A. P., Hawkley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: A single-item measure of positivity and negativity. *Cognition & Emotion*, *23*(3), 453-480.

Lombard, M. (1995). Direct responses to people on the screen: Television and personal space. *Communication Research*, *22*(3), 288–324.

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.

Massaro, D., Savazzi, F., Di Dio, C., Freedberg, D., Gallese, V., Gilli, G., & Marchetti, A. (2012). When art moves the eyes: a behavioral and eye-tracking study. *PloS one*, *7*(5), e37285.

Mather, G. (2013). *The psychology of visual art: Eye, brain and art*. Cambridge University Press.

McManus, I. C., Zhou, F. A., l'Anson, S., Waterfield, L., Stöver, K., & Cook, R. (2011). The psychometrics of photographic cropping: The influence of color, meaning, and expertise. *Perception*, *40*(3), 332-357.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's" up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology*, *93*(5), 699–710.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243-247.

Monk, S. (1935). *The sublime: A study of critical theories in XVIII-century England*. New York: Modern Language Association of America.

Nicolson, M. H. (1959). *Mountain gloom and mountain glory: The development of the aesthetics of the infinite*. Ithaca: Cornell University Press.

Pelowski, M., Hur, Y. J., Cotter, K. N., Ishizu, T., Christensen, A. P., Leder, H., & McManus, I. C. (2021). Quantifying the if, the when, and the what of the sublime: A survey and latent class analysis of incidence, emotions, and distinct varieties of personal sublime experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(2), 216–240.

Pelowski, M., Leder, H., Mitschke, V., Specker, E., Gerger, G., Tinio, P. P., ... & Husslein-Arco, A. (2018). Capturing aesthetic experiences with installation art: an empirical assessment of emotion, evaluations, and mobile eye tracking in Olafur Eliasson's "Baroque, Baroque!". *Frontiers in Psychology*, *9*, 1255.

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna integrated model of top-down and bottom-up processes in art perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80-125.

Pelowski, M., Specker, E., Gerger, G., Leder, H., & Weingarden, L. S. (2020). Do you feel like I do? A study of spontaneous and deliberate emotion sharing and understanding between artists and perceivers of installation art. *Psychology of Aesthetics, Creativity, and the Arts, 14*(3), 276–293.

Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. (2015). Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology*, *108*(6), 883–899.

Ward, P. I., Mosberger, N., Kistler, C., & Fischer, O. (1998). The relationship between popularity and body size in zoo animals. *Conservation Biology*, *12*(6), 1408-1411.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45-48.

Reymond, C., Pelowski, M., Opwis, K., Takala, T., & Mekler, E. D. (2020). Aesthetic Evaluation of Digitally Reproduced Art Images. *Frontiers in Psychology*, *11*.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1–21.

Seidel, A., & Prinz, J. (2018). Great works: A reciprocal relationship between spatial magnitudes and aesthetic judgment. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(1), 2–10.

Shiota, M. N., Keltner, D. J., & Mossman, A. (2007). The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and Emotion*, *21*(5), 944–963.

Silvera, D. H., Josephs, R. A., & Giesler, R. B. (2002). Bigger is better: The influence of physical size on aesthetic preference judgments. *Journal of Behavioral Decision Making*, *15*(3), 189-202.

Thouless, R. H. (1931). Phenomenal regression to the 'real' object I." *British Journal of Psychology*, *21*, 339.

Troscianko, T., Meese, T. S., & Hinde, S. (2012). Perception while watching movies: Effects of physical screen size and scene type. *i-Perception*, *3*(7), 414-425.

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

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., ... & Nadal, M. (2019). Preference for curvilinear contour in interior architectural spaces: Evidence from experts and nonexperts. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 110–116.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045.

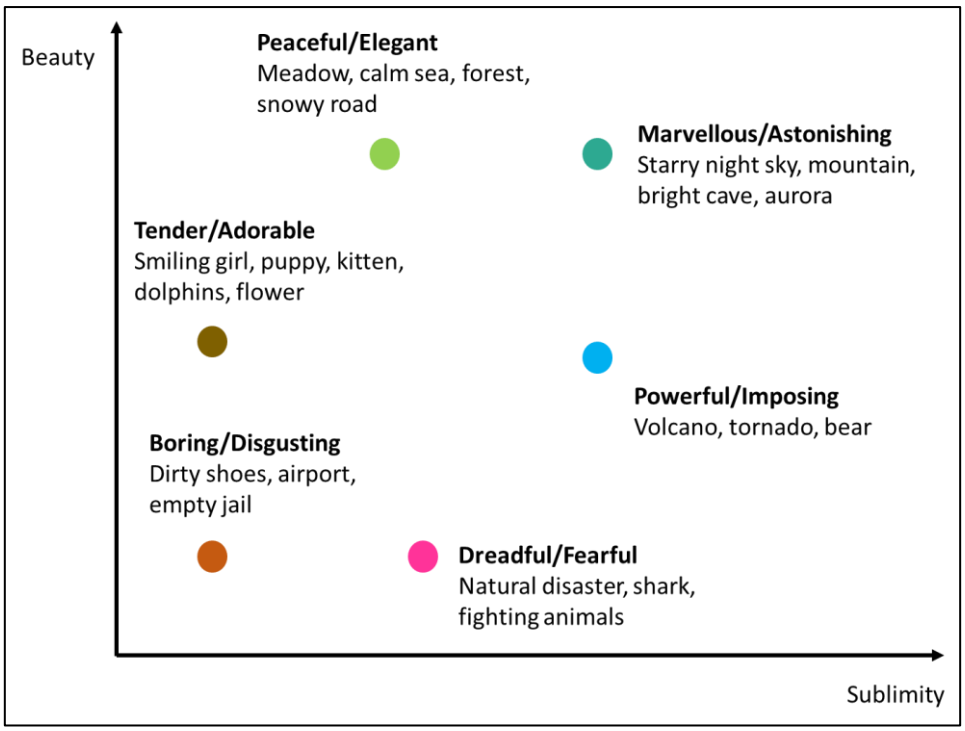

Figure 1. Conceptual figure of the *Aesthetic Hexagon*.

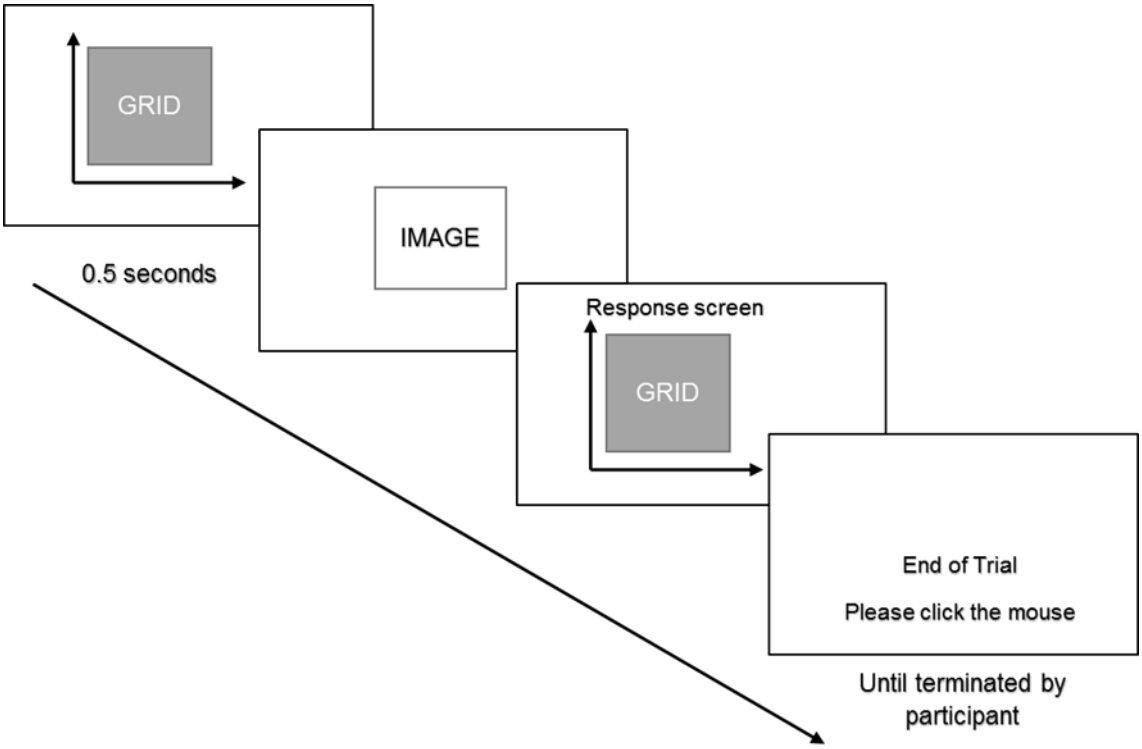

Figure 2. Schematic description of a trial progression.

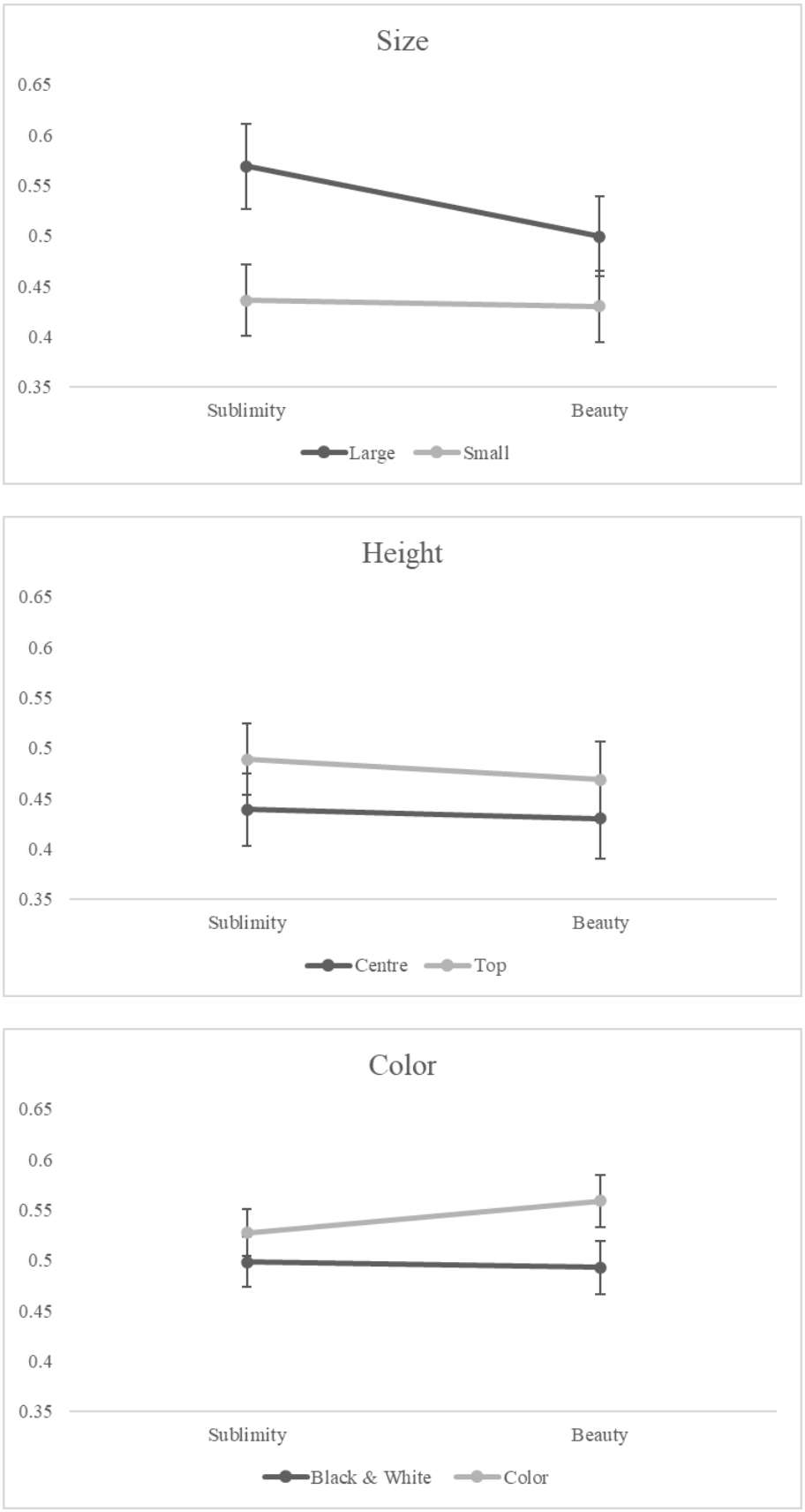

Figure 3. Size, height (above Study 1), and color effects (Study 2) visualization, with mean ±1SE.

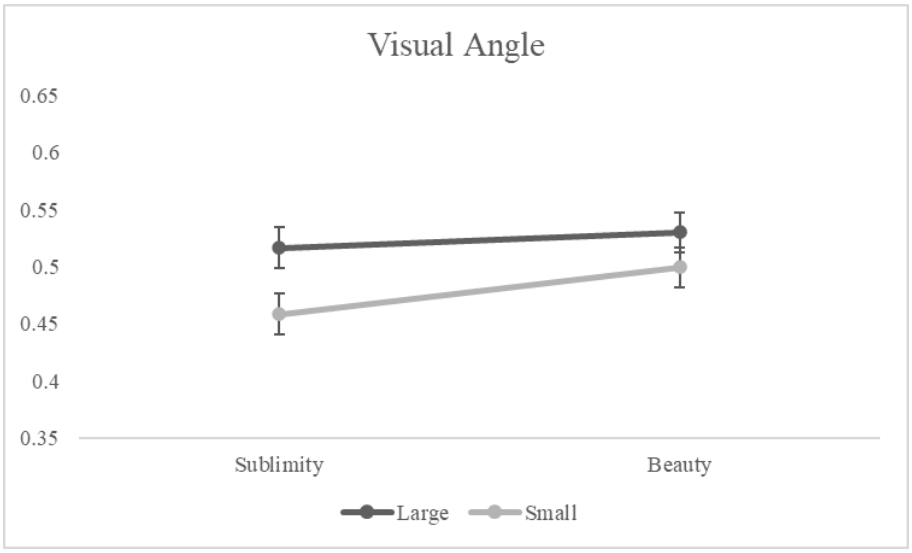

Figure 4. Visual angle effect (aggregated data analysis) visualization, with mean ±1SE.

Table 1a.

Descriptive statistics table for Study 1

		Sublimity rating	Beauty rating
Size Block	Large	0.56 (*SD* = 0.38)	0.49 (*SD* = 0.34)
	Small	0.43 (*SD* = 0.32)	0.43 (*SD* = 0.32)
Height Block	Top	0.44 (*SD* = 0.32)	0.43 (*SD* = 0.32)
	Centre	0.49 (*SD* = 0.36)	0.47 (*SD*= 0.34)

Note. Descriptive statistics represent raw mean values across all trials per condition.

Table 1b.

Size effect, inferential statistics table for Study 1

	df	*F*	*p*
Size	1, 49.74	53.45	**< .001**
Judgment Type	1, 68.62	0.73	.40
Size × Judgment Type	1, 81.53	19.51	**< .001**

Note. Bold shows inferential statistics that are significant at $p < .05$.

Table 1c.

Height effect, inferential statistics table for Study 1

	df	*F*	*p*
Height	1, 38.50	20.57	**< .001**
Judgment Type	1, 68.52	0.14	.71
Height × Judgment Type	1, 38.08	0.42	.52

Note. Bold shows inferential statistics that are significant at $p < .05$.

Table 2a.

Size and color effects, descriptive statistics table for Study 2

		Sublimity rating	Beauty rating
Large	Color	0.53 (*SD* = 0.31)	0.56 (*SD* = 0.30)
	Monochrome	0.52 (*SD* = 0.30)	0.51 (*SD* = 0.30)
Small	Color	0.51 (*SD* = 0.31)	0.54 (*SD* = 0.31)
	Monochrome	0.47 (*SD* = 0.30)	0.48 (*SD* = 0.28)

Note. Descriptive statistics represent raw mean values across all trials per condition.

Table 2b.

Size and color effects, inferential statistics table for Study 2

	df	*F*	*p*
Size	1, 85.70	19.24	**< 0.001**
Color	1, 41.80	14.33	**< 0.001**
Judgment Type	1, 38.10	1.11	.299
Size × Color	1, 3428.90	1.43	.231
Size × Judgment Type	1, 3458.00	1.22	.270
Color × Judgment Type	1, 3493.80	7.53	**.006**
Size × Color × Judgment Type	1, 3458.00	0.58	.446

Note. Bold shows inferential statistics that are significant at $p < .05$.

Table 3a.

Size effects, descriptive statistics table for Study 3

		Sublimity rating	Beauty rating
Large Dist.	Large V.A.	0.51 (*SD* = 0.31)	0.52 (*SD* = 0.29)
	Small V.A.	0.45 (*SD* = 0.30)	0.49 (*SD* = 0.28)
Small Dist.	Large V.A.	0.48 (*SD* = 0.31)	0.50 (*SD* = 0.28)
	Small V.A.	0.46 (*SD* = 0.30)	0.49 (*SD*= 0.28)

Note. "V.A." = Visual Angle. "Dist." = Distance. Descriptive statistics represent raw mean values across all trials per condition.

Table 3b.

Size effects, inferential statistics table for Study 3

	df	*F*	*p*
Visual Angle (V.A.)	1, 2613.81	14.97	**< .001**
Distance (Dist.)	1, 35.73	1.32	.26
Judgment Type	1, 120.80	1.17	.28
V.A. × Dist.	1, 2619.86	2.51	.11
V.A. × Judgment Type	1, 2673.53	5.79	**.016**
Dist. × Judgment Type	1, 2679.56	0.08	.77
V.A. × Dist. × Judgment Type	1, 2673.11	0.75	.39

Note. Bold shows inferential statistics that are significant at $p < .05$.

Table 4a.

Visual angle effect, descriptive statistics table for aggregated data analysis

	Sublimity rating	Beauty rating
Large	0.49 (*SD* = 0.31)	0.52 (*SD* = 0.29)
Small	0.45 (*SD* = 0.31)	0.47 (*SD* = 0.30)

Note. "V.A." = Visual Angle condition. Descriptive statistics represent raw mean values across all trials per condition.

Table 4b.

Visual angle effect, inferential statistics table for aggregated data analysis

	df	*F*	*p*
Visual Angle (V.A.)	1, 120.00	57.67	**< .001**
Judgment Type	1, 221.20	2.43	.12
V.A. × Judgment Type	1, 7587.80	18.01	**< .001**
Study (covariate)	3, 204.10	0.92	.43

Note. Bold shows inferential statistics that are significant at $p < .05$.

Pilot Studies

The Aesthetic Hexagon

In total, 96 participants (72 female, M_{age} = 23.19, SD_{age} = 9.17) were involved. The participants either consisted of University College London (UCL) students taking part for course credit or of individuals from the UCL subject pool taking part for financial compensation.

Of the 119 photographs, 52 photographs were sourced from the National Geographic website and 61 photographs were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997). IAPS is a set of emotion-evoking photographs standardized for their elicited valence, arousal, and dominance levels, and has been used in numerous emotion-evoking studies.

Not all participants saw the same set of photographs. However, all participants gave ratings of both sublimity and beauty for each photograph they saw.

Word Association Task

The following are the 112 words/phrases that all participants rated in their perceived association with sublimity and beauty: absorbed, active, admiration, adoring, afraid, agreeable, amazed, angry, anguished, arousal (vs. non-arousal), ascending (vs. descending), astonished, astonishing, at ease, awe-inspiring, beautiful, bemused, bored, calm, charmed, complex, conflict, confused, contemplative, controlling, curious, dark, delighted, delightful, delightful horror, desirous, disconcerted, disgusted, distanced, dominance (vs. submissiveness), dread, dreamy, elegant, elevating, energetic, enlightened, euphoric, exalting, existential awareness, fascinated, fearful, genius, grand, great, grief, happy, heightened perception, heroism, illuminated, immense, imposing, impressed, ineffable, innovative, intense (emotion-wise), interested, irritated, joyful, languorous, little, lively, loving, low (vs. high), magnanimous, majestic, marvellous, melancholic, monumental, mystical, nature, noble, novel, numinous, offended, openness and acceptance, orderly, passionate, peaceful, pleasant (vs. unpleasant), pleasure, power, presence, profoundness, psychologically distant, religion, sad, satisfied, sense of connectedness, sense of epiphany, sense of suspense, sensual, shaken, shocked, silent,

simple, smooth, softened, solemn, solemn sedateness, soothed, spiritual, stimulated, strength, surprised, tender, terror, touched.

Study 1

Manipulation Checks

Test-retest reliability, between-studies reliability, and between-participants agreement were assessed. These manipulation checks not only considered raw sublimity and beauty ratings, but also two derivative ratings. The degree to which an image evokes both sublimity and beauty was derived by adding up the ratings of sublimity and beauty (i.e. S+B). The degree to which an image evokes more sublimity than beauty was derived by subtracting up the ratings of beauty from sublimity (i.e. S-B).

Test-retest reliability. Since a photograph appeared twice per session, it was possible to measure the degree participants were consistent with their own ratings of photographic content over the two encounters. Although images appeared in two different presentation conditions, this approach enabled a rough measure of rating consistency. For each participant, a Pearson correlation was calculated between the evaluations. *r* values were averaged across participants via Fisher's Z transformations.[1] For interpretability, all calculations of correlations are reported in *r* values.

For the size manipulation block, the average test-retest reliabilities of sublimity and beauty were 0.83, and 0.87, respectively. S+B had a reliability of 0.89 and S-B had a reliability of 0.73. For the height manipulation block, the average test-retest reliabilities of sublimity and beauty were 0.89, and 0.88, respectively. S+B had a reliability of 0.90 and S-B had a reliability of 0.82. These calculations indicate a generally good level of test-retest reliability.

Between-studies reliability. 52 out of the 60 photographs were taken from pre-rated stimuli from the pilot studies. Therefore, it was possible to estimate to what degree the ratings from the current study are consistent with those from the pilot studies. Sublimity and beauty ratings were aggregated by item from the current and pilot studies respectively and were correlated. Both ratings were consistent between studies; sublimity, $r = 0.88$, $p < .001$, and beauty, $r = 0.91$, $p < .001$. The consistency was also found for the S+B, $r = 0.89$, $p < .001$, and S-B, $r = 0.94$, $p < .001$, ratings. This shows that the participants from the current study gave responses that are consistent with those of past

[1] $Z = \frac{1}{2}\ln\frac{1+r}{1-r}$

works, meaning that participants had similar ideas of the relative sublimity and beauty ratings of the stimuli.

Between-participants agreement. In estimating between-participants agreement, the "mean-minus-one" (MM1) correlation measure (Vessel, Maurer, Denker, & Starr, 2018) was used to determine the degree to which participants of the current study were similar to one another. As a measure of individual differences, the MM1 appears comparable to other known methods as pairwise correlation, inter-class correlation, and variance partitioning. Average ratings across repeated stimuli were computed for each participant. For a particular rating type, a set of correlations was then computed between an individual participant's ratings and the average ratings of all other participants. Using Fisher's Z transformations, these correlations were averaged together and backtransformed to an *r* value. The *r* value represented the MM1 score for that particular rating.

Subjected under the MM1, high levels of between-participants agreement as observed in sublimity, 0.86 (*SD* = 0.35), beauty, 0.85 (*SD* = 0.23), S+B, 0.87 (*SD* = 0.28), and S-B, 0.81 (*SD* = 0.25). These values are comparable to Vessel et al.'s (2018) highest reported MM1 values.

Study 2

Manipulation Checks

As in Study 1, the manipulation checks consisted of test-retest reliability, between-studies reliability, and between-participants agreement.

Test-retest reliability. Pearson correlations were calculated between the repeated evaluations of a certain image per participant (an image appeared twice in different size conditions). The average test-retest reliabilities of sublimity and beauty were 0.79, and 0.84, respectively. The average reliability scores were also acceptable for S+B, 0.85, and S-B, 0.73. This demonstrates that there was fairly good test-retest reliability in terms of sublimity and beauty ratings.

Between-studies reliability. All photographs rated in the current study had already been rated in the previous studies. Both sublimity and beauty ratings of the current work were highly consistent with those of past works, with $r = 0.88$, $p < .001$ and $r = 0.97$, $p < .001$ respectively. A similar set of between-study consistencies was found for S+B, $r = 0.95$, $p < .001$, and S-B, $r = 0.95$, $p < .001$.

Methodologically, the present study adopted a design where participants rated photographs after viewing each photograph for a fixed duration of 6 seconds. This was a contrast to past studies, where participants could view each photograph as long as they wished. In this context, the positive between-study reliability suggests that image viewing time does not affect one's relatively sublimity and beauty ratings.

Between-participants agreement. The "mean-minus-one" (MM1) correlation (Vessel et al., 2018) was calculated as a measure of between-participants agreement. Details of the method are available in Study 1. Subjected under the MM1, acceptable levels of between-participants agreement as observed in both sublimity, 0.70 (SD = 0.28), and beauty, 0.79 (SD = 0.35). Similar degrees of agreement were found for S+B, 0.75 (SD = 0.31), and S-B, 0.73 (SD = 0.36). These MM1 scores are considered high in the Vessel paper.

Parts of the results section of Study 2:

Table 1.

Inferential statistics table, for separate predictors.

	Predicting sublimity			Predicting beauty		
	df	*F*	*p*	*df*	*F*	*p*
Size	1, 40.68	20.52	< **.001**	1, 37.09	6.78	**.01**
Color	1, 114.53	2.07	.15	1, 46.08	20.60	< **.001**
Size × Color	1, 62.29	3.37	.07	1, 41.02	0.03	.87
(Covariate)	1, 2159.66	23.34	< **.001**	1, 2270.77	15.19	< **.001**

Note. Bold shows inferential statistics that are significant at $p < .05$.

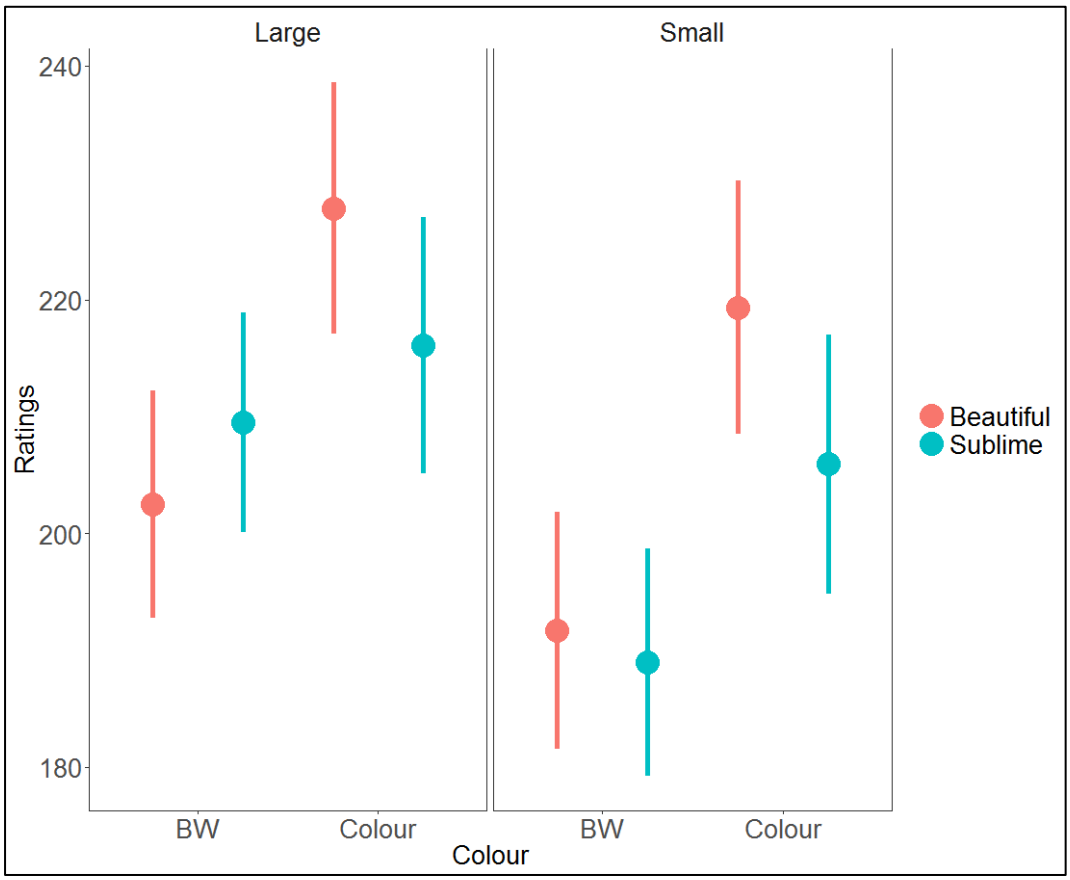

Figure 1. Color and size effects visualization, with mean ±1SE.

Study 3

Manipulation Checks

Because each participant looked at one image per session, test-retest reliability was not available. The 12 image contents from Study 2, averaged across all possible size, brightness, and contrast conditions, were rated consistently between studies. This was the case for sublimity, $r = 0.94$, $p < .001$, beauty, $r = 0.98$, $p < .001$, S+B, $r = 0.99$, $p < .001$, and S-B, $r = 0.97$, $p < .001$. Between-participants agreement measured as "mean-minus-one" (MM1) correlations (Vessel et al., 2018), too, was good for sublimity, 0.68 (SD = 0.35), beauty, 0.77 (SD = 0.26), S+B, 0.77 (SD = 0.28), and S-B, 0.63 (SD = 0.30).

Parts of the methodology section of Study 3:

Table 2.

Descriptive statistics for the brightness and contrast conditions.

	Mean Brightness Levels	Mean Contrast Levels
High Brightness Condition	0.48 (SD = 0.19)	0.20 (SD = 0.06)
Low Brightness Condition	0.15 (SD = 0.09)	0.19 (SD = 0.07)
High Contrast Condition	0.33 (SD = 0.16)	0.29 (SD = 0.06)
Low Contrast Condition	0.33 (SD = 0.16)	0.14 (SD = 0.05)

Note. Each calculation is based on the total set of 102 stimuli.

Parts of the results section of Study 3:

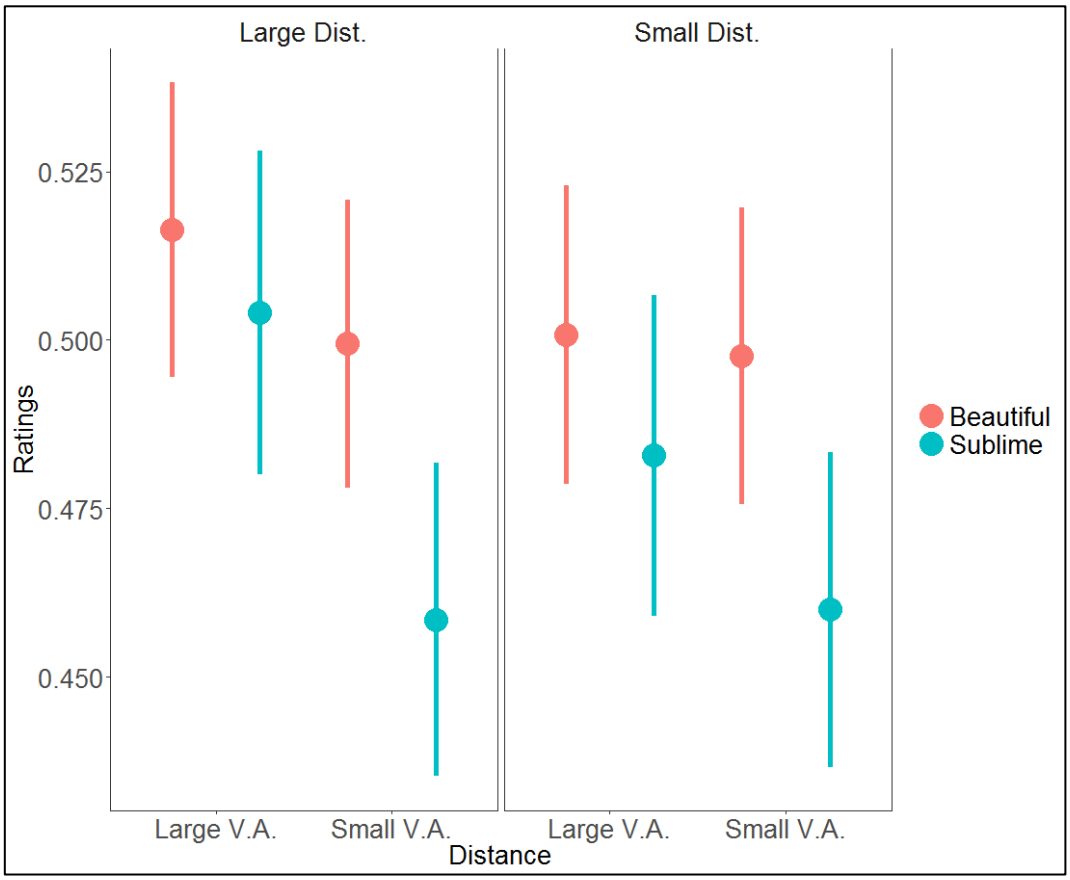

Figure 2. Size effects visualization, with mean ±1SE.

Table 3.

Size effects for separate predictors, inferential statistics table.

	Predicting sublimity			Predicting beauty		
	df	*F*	*p*	*df*	*F*	*p*
V.A.	1, 68.53	13.38	< **.001**	1, 80.59	0.72	.40
Dist.	1, 39.25	0.34	.56	1, 40.31	1.55	.22
V.A. × Dist.	1, 2639.16	2.40	.12	1, 2636.78	0.18	.67
(Covariate)	1, 2707.22	46.52	< **.001**	1, 2776.21	20.57	< **.001**

Note. "V.A." = Visual Angle. "Dist." = Distance. Bold shows inferential statistics that are significant at $p < .05$.

Table 4.

Brightness and contrast effects, descriptive statistics table.

	Sublimity rating	Beauty rating
High Brightness	0.47 (*SD* = 0.31)	0.50 (*SD* = 0.29)
Low Brightness	0.48 (*SD* = 0.30)	0.51 (*SD* = 0.28)
High Contrast	0.47 (*SD* = 0.30)	0.49 (*SD* = 0.28)
Low Contrast	0.48 (*SD* = 0.31)	0.51 (*SD* = 0.28)

Note. Descriptive statistics represent raw mean values across all trials per condition.

Table 5.

Brightness and contrast effects, inferential statistics table.

	df	*F*	*p*
Brightness/Contrast	3, 97.50	0.79	.50
Judgment Type	1, 5328.10	19.31	**< .001**
Brightness/Contrast × Judgment Type	3, 5328.10	0.31	.82

Note. Bold shows inferential statistics that are significant at $p < .05$.

Aggregated data analysis

The aggregated data analysis involved participants from Studies 1, 2, and 3, and an additional cross-modality study not reported in the present work. This additional study had thirty-nine participants (40 female, M age = 18.95, SD age = 1.26) and was approved by University College London's Research Ethics Committee. Across these participants, thirty-six pre-rated photographs representative of the *Aesthetic Hexagon* were rated on sublimity and beauty. All of these thirty-six photographs appeared in the three numbered empirical studies in the present paper.

The setting of this particular study was identical to that of Studies 1, 2, and 3. Images were presented 390cm away from the viewer, with visual angles of 21.77˚ x 38.70˚ (i.e. 150cm x 200cm).

While this additional study had cross-modality blocks in subsequent blocks, the present work's aggregated data analysis incorporates only the block where the participants rated the images only.

Linear Mixed Modelling (Sample code)

Below is a sample R code, where DV represents the dependent variable and IV represents the independent variable. The codes have been generated using the guidance of Judd et al.'s (2017) supplementary materials.

```
Model = lmer(DV ~ IV + (1+IV|Participant) + (1+IV|Stimulus) + (1|Participant:Stimulus), Data)
```

In the code, the "(1+IV|Participant)" indicates that there are differing baseline levels of the DV (the intercept, represented by "1") for each participant, as well as differing responses to the IV for each participant ("IV|Participant"). Likewise, the "(1+IV|Stimulus)" means that there are differing baseline levels of the DV for each stimulus, as well as differing responses to the IV for each stimulus. The "(1|Participant:Stimulus)" means that the model also considers the interaction between the random effects structures of participant and stimulus.