Standard Article

i-PERCEPTION

i-Perception
2026, Vol. 17(0) 1–31
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20416695261421103
journals.sagepub.com/home/ipe

Sage

Being K. Malevich: A hands-on approach to compositional preference

Doris I. Braun, Mara Hofmann, and Katja Doerschner

Abstract

Perceived visual balance is one of the critical factors that determine whether an artwork is regarded as aesthetically pleasing or not. This kind of balance seems to depend on the perceptual weight given to the major elements of a composition that can differ in size, position, shape, orientation, and colors within the pictorial field. We probe the relationship between perceived stability, dynamics, balance and aesthetic ratings of ten original artworks from the Suprematist art movement and 20 participant-arranged compositions with the elements taken from the artworks. For each artwork, 22 art-naïve participants arranged first all shape elements on the corresponding "canvas" of the paper version separately according to two tasks: create a stable and create a dynamic composition. After this, they transferred the final compositions to the computer and rated their digital versions of the stable and dynamic arrangements as well as the original Suprematists' artworks according to the degree of perceived balance, stability, and dynamic appearance and personal preference. Qualitatively, in "stable" compositions shape elements were arranged more symmetrically and often stacked on the lower part of the "canvas," whereas in "dynamic" compositions elements were often rotated with respect to the pictorial field and distributed unevenly or asymmetrically over larger areas. The relationship between perceived stability, dynamics, and liking varied with the complexity of the artwork and geometrical indices of balance did not correlate strongly with perceptual ratings and liking. Our analysis revealed differences in individual preference for either stable or dynamic arrangements.

Keywords

visual arts and aesthetics, perception of balance, stability and dynamics in compositions, interactive task, method of production, suprematism art movement, complexity, individual preferences

Received: May 13, 2025; accepted: January 20, 2026

How to Cite this Article

Braun, D. I., Hofmann, M., & Doerschner, K. (2026). Being K. Malevich: A hands-on approach to compositional preference. *i-Perception*, x(x), 1–36. https://doi.org/10.1177/20416695261421103

Justus-Liebig University, Giessen, Germany

Corresponding Author:
Katja Doerschner, Justus-Liebig University, Giessen, Germany.
Email: katja.doerschner@psychol.uni-giessen.de

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Introduction

An artwork, like a painting, drawing or collage is often created on a rectangular reference frame, that is, a canvas, sheet of paper, wooden plate or monitor. This is also how we encounter most of these artworks: Often nicely framed, hanging in rows along at the walls in museums or private rooms, offices, schools, or printed in books. When looking at an artwork, we have an instantaneous sense of its balancedness, that is, whether the composition is well composed with respect to its reference frame or background, or whether it "topples" or "flies out of the frame." Balance is one of the critical factors that determine whether we find an artwork aesthetically pleasing or not (Arnheim, 1954; Kandinsky, 1928; Locher et al., 1998; Ross, 1907); for a review see: Palmer et al. (2013). In fact, an unbalanced artwork becomes "incomprehensible" according to Arnheim (1954). However, what exactly is a "balanced" artwork and is the preference for balance and stability the same for everyone?

There is no simple answer to this question because the internal processes underlying aesthetic perception in humans are very complex and interactions between specific stimulus features and subject specific factors (age, education, sex etc.) that influence and determine aesthetic appreciation are manifold (Chatterjee & Vartanian, 2014; Fechner, 1871; 1876; Leder et al., 2004; Pearce et al., 2016; Specker et al., 2024). Several factors interactively modulate the impression of an overall balanced composition: Geometrical aspects like the size and shape of the pictorial field (canvas, paper, wall etc.), the painting style, the number, sizes, colors, structures and shapes of picture elements, their specific distribution with respect to symmetries, alignment, order, luminance and chromatic contrasts and the overall complexity (Arnheim, 1954; Belke et al., 2006; Corradi et al., 2019; Eisenman & Gellens, 1968; Fechner, 1871; Gartus & Leder, 2013; Hübner & Thömmes, 2019; Jacobsen & Höfel, 2002; Koenderink et al., 2018; Locher, 2003; Locher et al., 2001; McManus, 2005; McManus et al., 1985; Nadal et al., 2010; Specker et al., 2020; Van Geert & Wagemans, 2020). Also more semantic aspects such as the topic, time period, uniqueness of the artwork play a role in addition to the artwork-inherent factors, the specific circumstances, like where and when the artwork is seen, the information provided or known and effects of repeated exposure (Leder, 2002; Leder et al., 2004; Tinio & Leder, 2009). Finally, also the observer brings his/her own comprehension, expertise, prior experiences, current state and emotions to the judgment of a composition (Arnheim, 1954; Chatterjee & Vartanian, 2014; Leder et al., 2012; Marković, 2011; Palmer et al., 2013; Pearce et al., 2016; Specker et al., 2017).

According to Arnheim (2009) the geometric center of a pictorial field "is also the center of the composition, in the sense that all the perceived weights of shapes and colors arranged by the painter balance around the middle" (p. 66). These weights exert perceptual forces and when these forces act on an artwork's center of gravity, they balance each other out such that an equilibrium, that is, a perceptual balance has been achieved (Arnheim, 1954). But is this true and is this kind of balance perceived by all observers? A composition that looks balanced to me may look unbalanced to you. Given the complex interactions of very different factors, how can one even hope to break any way towards understanding the visual perception of balance in artworks? One possibility is to simplify the stimuli and to start by studying the perception and preference for different arrangements of single elements. For example, Ross (1907) discussed in his book, how the shape, position, orientation, and length of a single line on a canvas or paper influenced aesthetic impressions (also see Puffer, 1903). McManus et al. (1985) found that observers generally indicated pictures were balanced near the center of the frame, that balance judgments were based upon the position of an object and influenced much by size or color and that the whole information present in the pictorial field was integrated. Locher et al. (1998) used sets of similar geometric shapes of different sizes, for example, black circles, squares, leaves etc. and asked their participants to arrange them on a white quadratic background ("canvas") until the arrangements appeared "interesting and pleasant" to them. They found that the structural center, that is, the point "about which there is an equal distribution of the physical weight associated with its components" of the resulting designs often corresponded to the geometric center of the "canvas". Therefore, participants appeared to have used an imaginary horizontal and vertical grid while creating their compositions. Their results support the idea that "perceptual balance" might be thought of it as "mechanical balance," where each visual element has a certain perceptual weight depending on its shape, color, and size (Arnheim, 1954).

Quantifying Balance

Although being criticized for being a too simplistic notion (Abeln et al., 2016; McManus et al., 2011), in recent years, the "mechanical" conception of perceptual balance has inspired the development of quantitative indices of balance that can be computed on digitized images of simple compositions of geometric shapes. For example, (Wilson & Chatterjee, 2005) introduced the "Assessment of preference for balance" (APB) index. They validated this index by comparing APB scores and human balance ratings for 130 images consisting of white squares on which seven black elements of the same shapes (circles, squares, or hexagons) of varying sizes were distributed differently. For the index an image was divided into two equal parts along the two principal axes, two diagonal axes, and into equal inner and outer areas. Pixels on each side with respect to each axis were counted and the relative measure of symmetry was calculated. Then the average of the eight different measures of symmetry, all based on the comparison of pixel distributions of the black shapes with respect to geometric divisions, was calculated. This average yielded the APB index, the objective balance score ranging from 0% (perfect balance) to 100% (no balance).

Hübner and Fillinger (2016) developed, based on Arnheim's theory (1954), an alternative measure, the so-called "Deviation of the Center of Mass" (DCM) that quantified the deviation of the perceptual center of "mass" from the geometric center of the image. Like Wilson and Chatterjee (2005) they created pictures with seven black elements (circles and hexagons) of different sizes, that were located at random positions. In three different experiments participants had to rate the balance and symmetry and their preference for these pictures. Similarly, to the APB method, each black pixel of each shape element on the "canvas" received a "mass" and their sum was weighted by its distance with respect to a reference point in the horizontal and vertical direction, respectively. Finally, the combined horizontal and vertical distances of the visual "mass" to the geometrical center of the image are computed as the DCM (see Hübner and Fillinger (2016) for equations). The authors showed that the DCM and APB predicted balance and liking rating similarly well, with APB accounting for 62% and DCM for 68% of the variance in balance ratings (Hübner & Fillinger, 2016). However, in a subsequent study (Hübner & Fillinger, 2019) pointed out that these strong predictions might be due to the specific set of test images, that is, the use of homogeneous shape elements with the same luminance level and that they may not generalize well to more complex stimuli. For example, both measures, the APB and DCM, failed to predict balance ratings for Japanese calligraphy (Fillinger & Hübner, 2020b; Gershoni & Hochstein, 2011). This was also found in a subsequent study with stimulus arrangements of varying complexity of the visual aesthetic sensitivity test (Götz et al., 1979); in which measures of APB and DCM did not correlate with balance ratings of participants, except for single element stimuli (Hübner & Fillinger, 2019). The authors proposed that it might be possible that participants "apply concepts of perceptual balance that are not reflected by the APB and DCM measures" (Hübner & Fillinger, 2019), such as "a sense of stability." They conclude that how participants interpret perceptual balance depends on the type of picture, for example, for single-element pictures participants use a concept of "mechanical balance" as captured by the DCM (and APB), whereas for multiple-element and dynamic-pattern pictures balance is interpreted as "gravitational stability" (Hübner & Fillinger, 2019).

Balance and Aesthetic Experience

In order to better understand the effects of perceived and image-computed balance on aesthetic experience, the authors of the studies above (Hübner & Fillinger, 2019) also assessed how much an picture was liked, and related mean ratings of liking to the mean rating for balance and quantitative measures of the APB and DCM scores. A subgroup of participants rated, instead of balance, the perceived stability. Their results suggest that aesthetic appreciation of a picture is not just dependent on perceived balance but on several components, such as the (perceived) stability, complexity (here, number of elements; but also see Nath et al. (2023) for additional measures of complexity) and the perceived dynamics of the composition. In fact, overall, ratings and indices of balance did not correlate with liking, instead liking strongly depended on the complexity of the picture: The higher the complexity of a picture the more participants liked it (Hübner & Fillinger, 2019). Overall, an interesting pattern emerged: For single element stimuli, liking was unrelated to perceived balance and stability, but positively correlated with complexity, whereas for multiple-element stimuli liking was positively correlated with perceived balance and stability, but unrelated to complexity, and finally, for dynamic distributions of elements, liking was positively correlated with perceived balance and complexity, but unrelated to stability. The authors argue that these different patterns of ratings might be explained by the fact that, depending on the picture type, participant use different criteria for their ratings. For example, compositions that imply motion may be evaluated with different criteria for balance and stability than those that do not. Therefore, they proposed that two types of instability exist: One type of instability is associated with gravity and is not liked, whereas the other type is associated with movement, which is liked (Fillinger & Hübner, 2020a).

To test this idea Fillinger and Hübner (2020a) systematically assessed the complex interplay between liking, perceived balance, stability and implied motion as well as the two quantitative measures of balance, APM, and DCM. They used two different sets of stimuli: For experiment 1, they created 32 compositions inspired by the artists K. Malevich (1879–1835) containing three black rectangles and three small colored elements in different orientations, for experiment 2 they selected 100 artworks by Karl Otto Götz (1914–2017) and presented them as gray-level pictures. For the multiple-element stimuli in experiment 1 they found a strong positive correlation between balance and stability ratings: The balanced and gravitationally stable compositions were liked by participants, but not the more dynamic ones. Moreover, DMC scores correlated only with the balance-, but not with stability ratings, which Fillinger and Hübner (2020a) interpret such that balance ratings might reflect a mixture of mechanical balance and gravitational stability. For the more complex, textured, and dynamic images of artworks in experiment 2, instability (dynamics) was liked, and no correlation between liking and balance ratings were found. Also, none of the perceptual ratings for these images correlated with DCM (and APB) scores. Fillinger and Hübner (2020a) interpreted the differences between the results of the two experiments as support for the idea that perceived instability reduced the aesthetic appreciation of a composition unless it implies movement. However, given that the stimuli in experiment 1 and 2 were of vastly different quality—one consisting of a few geometric similar shapes and the other one of complex paintings—comparing them directly might be problematic. Moreover, there is an implication that compositions consisting of a few geometrical elements do not imply any motion, which isn't necessarily true. As an example, consider "White on White" by K. Malevich shown in Figure 1 (right), a single tilted white square on white background. This composition consists of just a single geometric element defined by hardly any luminance contrast, yet it appears as quite dynamic (at least to us).

Given that the relationship between perceived balance, implied motion and liking is complex and also dependent on the image type, it could be helpful to keep some of these parameters fixed while varying others. Here we attempt this by asking participants to arrange their own compositions that

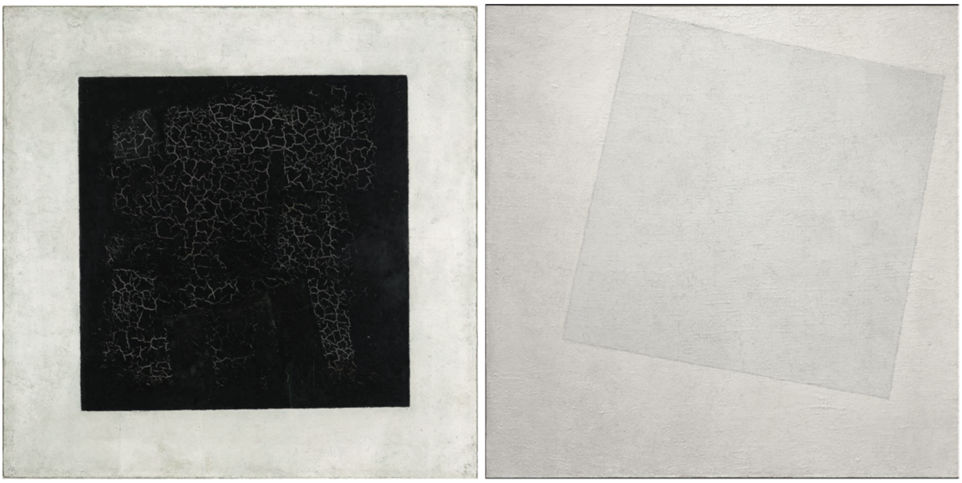


Figure 1. The black and the white square by K. Malevich. Left: Black square by Malevich (1915), oil on canvas (79,5 × 79,5 cm), Trejakow-Galerie Moskau. Malevich's famous painting *Black Square* is considered as a major landmark in the history of abstract art and Malevich painted four versions of it between 1915 and the early 1930s. At the *0.10* exhibition in 1915 it hung underneath the ceiling across one corner of the exhibition room, which was the traditional placement of religious icons in Russia. A close inspection reveals that the square is not a real square since its horizontal outlines are not parallel. (source of image and info: https://www.tate.org.uk/art/artists/kazimir-malevich-1561/five-ways-look-malevichs-black-square) Right: White on White by Malevich (1918), oil on canvas (79.4 × 79.4), MoMA, New York. In this painting a barely visible off-white square is superimposed on an off-white ground (image source: https://www.moma.org/collection/works/80385).

are either stable or dynamic with predetermined sets of few geometric shape elements. Elements were taken from original Suprematist artworks. Suprematism is an art movement founded by the Ukrainian artist and art theorist Kazimir Severinovich Malevich together with a group of mainly Russian artists, with the aim creating a non-objective "pure artistic feeling." These artworks usually consist of few basic geometric shape elements like circles, lines, rectangles and squares painted in a limited range of colors (see Figures 1 and 2), that made them particularly suitable for our purposes (also see Information A1 for background on the suprematist art movement).

Arrangement tasks are intuitive: They don't require the participant to have artistic drawing or painting abilities, and one can easily manipulate, for example, the complexity of compositions by varying the number of elements. The results of such interactive tasks are quite insightful for understanding what composition aspect elicits a sense of pleasure or liking and how this experience is linked to the arrangement of elements. Fechner, in his book "Zur experimentellen Aesthetik" (1871, pp. 48–49), introduced the method of production as one of three possible ways to study aesthetic experiences. With this method, participants actively produce arrangements that "please" them, yielding subjective arrangements whose properties can then be analyzed subsequently. Despite this potential, most studies of empirical aesthetics are done with the method of choice, that is, with a two-alternative forced choice task or a rating paradigm with a Likert scale since testing is fast and the analysis of the collected data is easier. The method of production is still rarely used (Westphal-Fitch, 2019) although it is much closer to the artistic process and allows more individual freedom (Braun & Doerschner, 2019). Recent examples of how the method of production is used in studying aesthetic preferences, can be found in Westphal-Fitch et al. (2012) or Westphal-Fitch et al. (2013), where they introduced an interactive pattern-generation framework for participant to make their own pattern arrangements. In our former study, we asked our participants to adjust the color

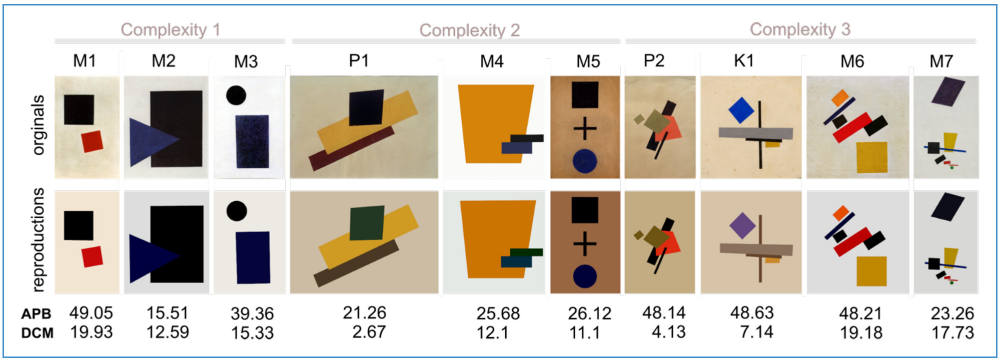


Figure 2. Digital images of the original Suprematist artworks of Malevič (M), Popowa (P), Khidekel (K). Top row shows the originals from left to right: (1)*M1*: "Black Square and Red Square," (2) *M2*: "Suprematism with Blue Triangle and Black Square," (3) *M3*: "Suprematism with rectangle and circle 1915," (4) *P1*: " Draft for embroidery for the collective Werbowka," (5) *M4*: " Suprematism 1915," (6) *M5*: "Suprematism with square, cross and circle," (7) *P2*: "Draft for embroidery number 36 for the collective Werbowka," (8) *K1*: "Suprematist Composition with blue square," (9) *M6*: "Frame Class," (10) *M7*: "Painterly realism of a soccer player." In this figure the images of the artworks are scaled to have the same height. The bottom row shows the manually reconstructed artworks in PsychoPy (Peirce et al., 2019). Colors in reconstructed artworks in this figure may slightly deviate from the actual color appearance on the experimental screen. Below each artwork we report the corresponding APB and DCM of original Suprematist artworks. APB and DCM were computed according to Wilson and Chatterjee (2005) and (Hübner & Fillinger, 2016) respectively. For APB and DCM lower numbers indicate higher balance (possible range: 0% to 100%). APB = assessment of preference for balance; DCM = deviation of the center of mass.

of a single gray shape within an abstract painting presented on a monitor according to their taste (Braun & Doerschner, 2019). However, the method of production also comes with its challenges. For example, McManus and Gesiak (2020) using interactive Mondrian patterns showed that too many degrees of freedom in creating compositions can be overwhelming for participants. Also, Fechner noted that the production task is best to be used with simple stimuli, for example, a few geometric shapes (Fechner, 1871). Recent work by Hübner (2025) suggests that production and evaluation (rating) tasks yield different preference patterns for aspects of a composition.

Here, we asked participants to create a stable and a dynamic composition for arrangements containing different numbers of shape elements (see Figure 2). After this, participants rated their own compositions as well as the original artwork on how stable, dynamic, and balanced it appeared, and also, indicated how much they liked each composition. The goal was three-fold: (1) To assess art naïve participants intuition about stability and dynamics in a visual composition, that is, whether there were any systematic patterns in the arrangements under each composition task (2) to assess to what extend perceived dynamics and stability correlated with the participants concept of visual "balance" and their liking, and how these relationships varied as a function of visual complexity. (3) To formally assess how well geometrical aspects of the compositions and established indices of balance—APB (Wilson & Chatterjee, 2005) and DCM (Hübner & Fillinger, 2016)—related to the subjective experience of self-arranged and original compositions.

Methods

Stimuli

Stimuli were constructed from ten Suprematist artworks by Malevič (M), Popowa (P), and Khidekel (K). The upper row of Figure 2 shows the original artworks; their high-quality digital reproductions were retrieved from wikiart.org (*M*1, *M*2), meisterdrucke.de (*M*5, *M*7, *P*1, *P*2), kunst-fuer-alle.de (*M*3), reproarte.com (*M*4), germanposters.de (*M*6), artsandculture.google.com (*K*1). For our study compositions had to fulfill the following criteria: (1) Each artwork had to consist of 2–8 geometrically and chromatically distinct and clearly separable shapes, that were placed entirely within the artwork boundary (i.e., no cut-off elements). Note, that we will refer to an artwork or its set of shape elements taken from a specific artwork either generally as "stimulus," or by its shorthand, for example, *M*2 (see Figure 2). For more information about perception and Malevich's art and especially the painting "Black Square and Red Square" (1915) our stimulus *M*1, see Sarami et al. (2021).

Shape elements and the background of each artwork were carefully measured and all 10 artworks were re-constructed in Python using PsychoPy (Peirce et al., 2019) Routines (Figure 2 bottom row). We approximated the colors of all compositions as close as possible to their appearance in our specific digital reproductions of the artworks. This allowed us to access each element in a given composition and control its position in the experiment. For each artwork we printed all individual shape elements and the "canvas" on thick A4 paper and cut them out. This allowed our participants to get familiarized with the compositional tasks and made it easy for them to explore the effect of positions of the given shapes within the pictorial space. We are aware that colors of digital reproductions and paper-cutouts deviated somewhat from the "analog" original color additional to the sizes. However, this problem can't be solved and we hope that these facts are not too relevant to our research question. Geometrical measurements and the experimental code can be obtained from the authors upon request.

Participants

Twenty-one volunteer participants (13 female, age range 21–33 years, mean = 24.5, *SD* = 3.23,) were recruited using the Justus-Liebig-University's mailing list. Data was acquired within two weeks. Nineteen participants were students at the JLU, two others had completed vocational training. Upon arrival in the laboratory, participants received general instructions about the design of the study. Then they signed the consent form, and filled in a short general presurvey inquiring about demographic data, favorite and least liked colors, general interest in art and abstract art. On average, participants indicated to have little to intermediate interest in visual arts and none of them studied or was trained in art. All participants were right-handed, had normal or corrected-to-normal visual acuity and normal color vision as tested before with the Ishihara test for color deficiency (Ishihara, 2018). The study was approved by the local ethics committee (LEK FB06) and was conducted in accordance with ethical standards put forward in the Declaration of Helsinki.

Experimental Tasks and Procedure

The study consisted of two experimental parts: In *part 1*, participants were asked to arrange the elements of each artwork (A) to achieve a dynamic (dynamic/energetic/moving/lively) composition and (B) to achieve a stable (stable/static/calm) composition. The order of A and B was randomized across participants. *Part 2* consisted of four sub-experiments: Participants were asked to rate their 20 self-arranged and the ten original compositions by moving a horizontal slider on how balanced (completely unbalanced—very balanced), likable (completely dislike—like it very much), stable (completely unstable—very stable) and dynamic (completely static—very dynamic) they appeared to them. Numerically, slider values ranged for every task between 0 and 1, where 0 referred to the absence of an attribute or dislike, and 1, conversely to the full presence of an attribute or much liking. The order of the four rating experiments was kept the same for all participants, but the 30 stimuli (10 stable compositions, 10 dynamic compositions, 10 original artworks) in each rating experiment were presented in a randomized order.

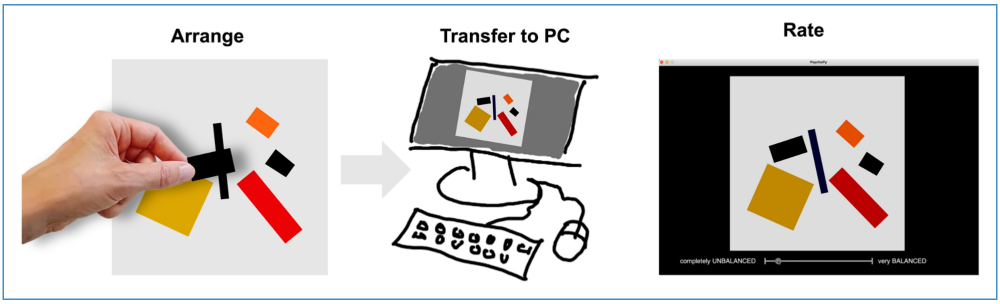


Figure 3. Experimental tasks. Left: Participants first completed the composition task with cut-outs for a given artwork. Middle: When they were content with their design, they transferred their composition onto a PC using a custom software written by us in PsychoPy (Peirce et al., 2019). Right: After doing this for all artworks and for both compositional tasks (balanced, dynamic) participants rated their own and original Suprematist artworks in four rating experiments. In a randomized order, participants rated each stimulus on how balanced, dynamic, stable, and likable it appeared to them. See the main text in the methods for more details.

Part 1 started with a short practice session. Participants were asked to arrange 3 shapes (these were not taken from any of the experimental stimuli) on an A4 paper until they achieved a satisfactory (to them) stable or dynamic composition. Then the experimental part 1 started. On every trial, participants arranged the elements for one randomly chosen artwork according to the given task first with the cut-out paper stimuli on the corresponding paper backgrounds. When they were content with their arrangements of the cut-outs, participants transferred them to the computer using a custom-written PsychoPy (Peirce et al., 2019) Program. Shape elements of a given composition were displayed on the left side of a rectangular shape representing the "canvas". The program allowed participants to modify position, rotational angle and depth order of each element by using the computer mouse and the arrow keys on a regular keyboard (see Figure 3 for illustration of the task). Participants locked in their composition and advanced to the next one by pressing the spacebar. This procedure was the same for all ten set of stimuli in the "dynamic" and "stable" tasks, yielding 20 compositions from every participant.

In part 2, the 20 compositions of each participant were randomly intermixed with the 10 computer re-constructed "originals" and presented in four experimental blocks, each with a different rating question (balance, preference, stability, dynamics). In each block, participants had to indicate their rating on an attribute for each of the 30 stimuli, for example, the perceived stability by moving a slider (Figure 3, right).

Upon completing both experimental parts participants filled out an exit survey asking for specific feedback concerning the experiments (difficulties, problems) and individual artworks. The majority of participants enjoyed the tasks, they reported to have used "gut feeling" when completing the compositions. About half of the participants reported the "stable" composition task to be easier, because it was perceived to be "unambiguous" and "intuitive" what it meant. Several participants reported the dynamic composition task to be easier when using multiple elements, than when using a few. Overall, no major problems were reported.

Apparatus and Software

During the experiments, a participant sat at a table in an illuminated room (daylight) when arranging the individual compositions and the rating task. Compositions were transferred to a computer using a custom-written experimental software that uses PsychoPy (Peirce et al., 2019) routines. The computer monitor (Display ++, LCD, Cambridge Research Systems Ltd), was linearized and driven at a 120-Hz refresh rate. At a viewing distance of 90 cm, the active screen area subtended 42.5 degrees visual angle horizontally and 24.45 degrees vertically and with a spatial resolution of 1920×1080 pixels 45 pixels were presented per degree visual angle. Each artwork (canvas and elements) was proportionally scaled such that it would take up 80% of the vertical space on the monitor. The position of the observer's eye corresponded approximately to the height of the screen center. The software for the rating experiments in part 2 was also written by us using PsychoPy (Peirce et al., 2019) routines. Stimuli appeared in the same screen position as in part 1.

Analysis

To test whether art-naïve participants can create compositions that are dynamic or stable, we assess in a 3-way analysis of variance (ANOVA) whether participant ratings varied systematically with their stimulus arrangement task (stable, dynamic (2 levels), artwork (10 levels) and perceptual rating question (stability, dynamics, balance i.e., 3 levels).

We computed DCM and APB scores for all stimuli, the 20 self-arranged compositions of each participant and the ten "original" Suprematist artworks, according to the formulas provided in Hübner and Fillinger (2016) and Wilson and Chatterjee (2005), respectively. We converted our stimuli to binary images, the background or canvas was set to white (1,1,1) and any pixel belonging to a shape element as black (0,0,0). We also repeated this computation with a weighted-by-pixel-intensity-level approach (Thömmes & Hübner, 2018), however, results did not deviate much. We verified our computations of APB and DCM scores by comparing our values to those provided in Hübner and Fillinger (2016) for their Figure 1, bottom row. We categorized original artworks according to three levels of complexity. As a very rough proxy for complexity we simply counted the number of shapes in a given composition (as in Hübner & Fillinger, 2019; Wilson & Chatterjee, 2005), however complexity is arguably much more than that, for example, see Nath et al. (2023). A low-level complexity artwork contained up to two shape elements ($M1$, $M2$, $M3$), a medium level up to four elements ($P1$, $M4$, $M5$) and a high level up to ten elements ($P2$, $K1$, $M6$, $M7$). For all stimuli, we computed DCM and APB indices as outlined above, and we also computed separately, the DCM constituents: That is, the horizontal deviation of the center of mass (HDCM) and vertical deviation of the center of mass (VDCM) distances of the visual "mass" to the geometrical center of the image. For each complexity level, we correlated all geometric indices with overall liking and perceptual ratings of our participants, examined the data for consistent differences in individual preference patterns by further analyzing ratings ("like") for dynamic and stable self-arranged compositions, as well as the, on average, most and least dynamic—and most and least stable rated original artworks. All rating and composition data is available at https://doi.org/10.5281/zenodo.18258565.

Results

Participants can Create Dynamic and Stable Compositions

Figure 4 depicts the 20 compositions of each participant. Most participants followed the instructions and were able to create stable and dynamic compositions. For the stable composition task, participants chose a similar main orientation for the shape elements and often stacked them solidly on top of each other in the lower part of the "canvas". For the dynamic composition task, they often rotated shape elements with respect to the "pictorial frame" and preferred more wide, uneven or asymmetric distributions. We can quantify these observations by measuring, for each artwork, how much the orientation of its elements deviated from a cardinal direction (i.e., horizontal/vertical, whichever was closer). On average, deviations were larger for compositions in the dynamic task (mean =

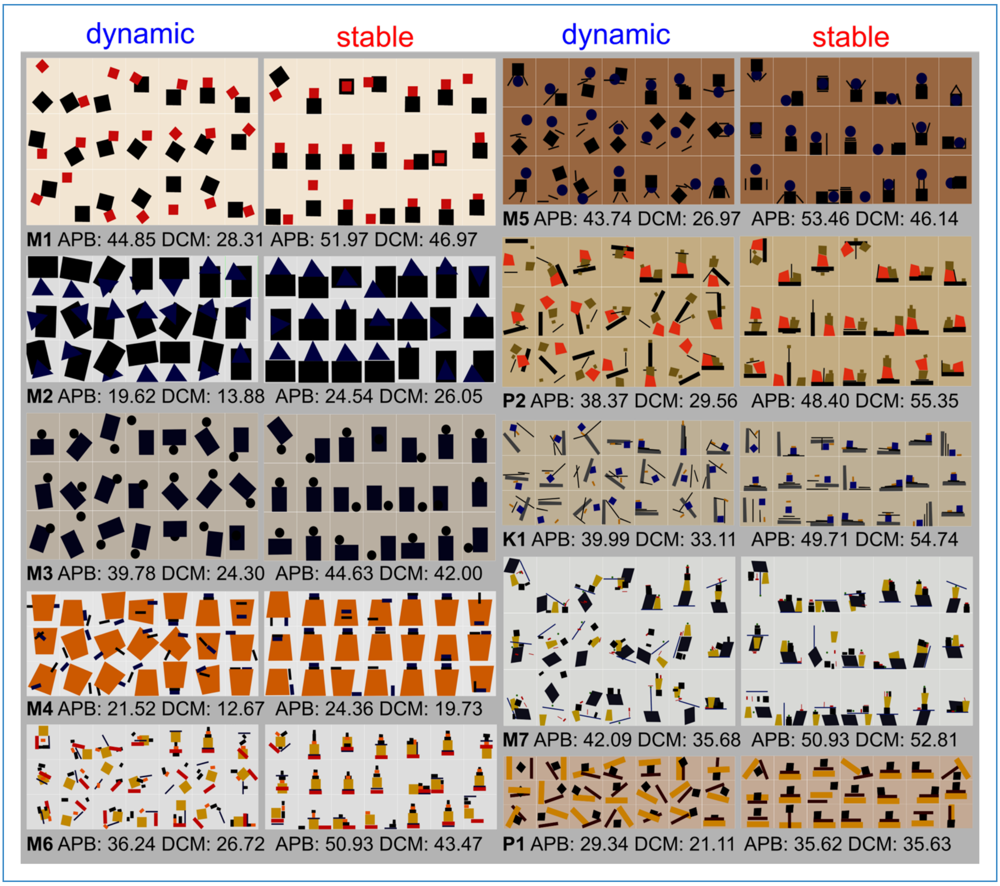


Figure 4. Participants' dynamic and stable compositions with shape elements taken from 10 Suprematist artworks. For each artwork (see Figure 2) all compositions of the 21 participants are sorted according to the two tasks as indicated above each column. The majority of participants arranged the positions, orientations, and distributions of shapes differently when they designed a dynamic and a stable composition. Participants were quite creative, especially for the dynamic compositions. In rare cases, a participant did not seem to follow the given instruction, for example, participant no. 1, stable compositions. We had divided the black cross of artwork *M5* (see Figure 2) into two equal bars, which inspired some participants to create human-like figures with two arms showing a more dynamic or a more stable "behavior." Below each set of compositions average APB and DCM values are provided (lower numbers indicate a higher balance). APB = assessment of preference for balance; DCM = deviation of the center of mass.

14.44 degrees) than in the stable task (mean = 2.68). This difference was significant, and a 2 (composition task: Stable, dynamic)×10 (artworks) ANOVA yielded a significant main effect of composition task $F(1{,}400) = 203.23$, $p < .0001$. There was also a small but significant main effect of artwork $F(9{,}400) = 2.04$, $p = .034$, but no significant interaction. Figure 5 shows the average deviation of compositional elements for each artwork, for compositions in the stable task (red) the dynamic task (blue) and original compositions (gray).

Below the compositions in Figure 4, average (across 21 compositions) APB and DCM indices are provided (as well as in Table A1 see Appendix). Overall, both DCM and APB values, tended to be significantly smaller for dynamic—than stable compositions ($DCM_{Dynamic} = 24.96$, $SD = 15.04$; $DCM_{Static} = 41.21$, $SD = 20.47$; $t_{DCM}(418) = -9.267$, $p < .0001$ $APB_{Dynamic} = 35.72$, $SD = 11.69$;

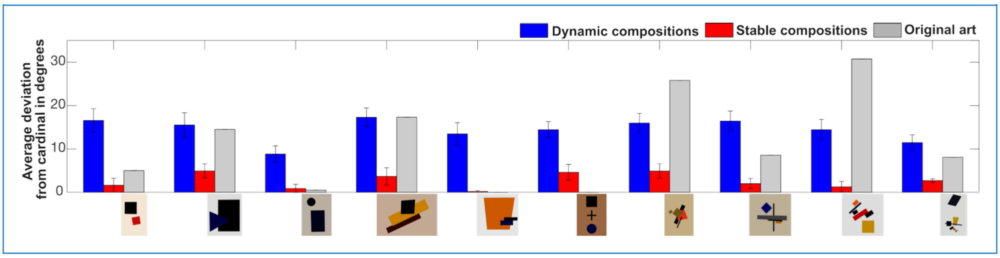


Figure 5. Average deviation of compositional elements from cardinal direction. For each artwork, we computed how much each compositional element deviated from the horizontal or vertical orientation (whichever was smaller) in degrees. These differences were averaged across all elements (in an artwork). And the averaged differences were averages across 21 participants. Note, that the shape of an element was disregarded in this analysis, for example, whether the shape was a circle or a square or a rectangle. Blue bars show the average deviation for dynamic-, red bars, for stable-, and gray bars for the original compositions. Error bars are one standard error of the mean. Overall, we can see that the deviation from cardinal was much larger in dynamic-, than in stable compositions. Original artwork varied in terms of their magnitude of deviation from cardinal orientations.

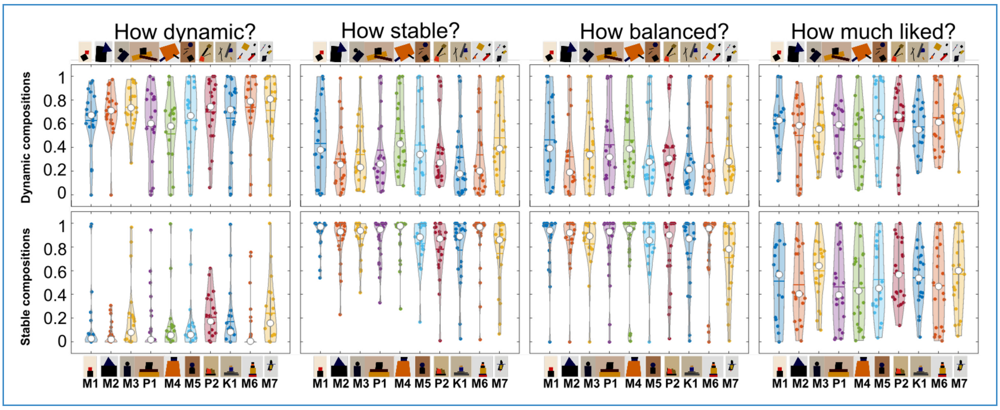
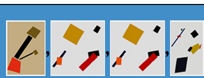
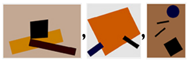
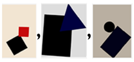


Figure 6. Rating data for participants' compositions for all ten artworks in two composition tasks. In the violin plots of the rating data individual participants data are presented by colored dots, white dots denote the median, and solid horizontal lines the mean. The shape of the violins helps to assess how the rating data is distributed. Compositions using the shape of individual artworks are represented along the *x*-axis by their shorthand as well as a representative participant composition. The *y*-axis indicates the rating strength for a given attribute/task (see Methods).

$APB_{Static} = 42.83$, $SD = 12.14$; $t_{APB}(418) = -6.106$, $p < .0001$). In the Appendix, Figure A1 shows the corresponding average HDCM and VDCM for dynamic and stable compositions.

We next formally assessed whether participants solved the two composition—and rating tasks in a reasonable way. We expected distinct rating patterns, for example, dynamic compositions should be rated as more dynamic and stable compositions as more stable. Overall, this appeared to be the case, as shown in Figure 6: On average, dynamic compositions were rated as more dynamic ($M = 0.659$, $SD = 0.283$) and less stable ($M = 0.393$, $SD = 0.317$), whereas stable compositions were rated as less dynamic ($M = 0.160$, $SD = 0.241$) and more stable ($M = 0.836$, $SD = 0.222$). A 2 (arrangement task: Stable, dynamic)×2 (rating question: Dynamic, stable) ANOVA yielded a significant main effect of rating question $F(1,836) = 122.26$, $p < .0001$, no main effect of arrangement

Table 1. Correlations between APB and DCM scores and participants' ratings for stable and dynamic compositions sorted by three levels of complexity.

	Stability	Balance	Liking	APB	DCM	VDCM
	High Complexity: *P2, K1, M6, M7*: 					
	Stability	Balance	Liking	APB	DCM	VDCM
Dynamics	**−0.6931*****	**−0.4876*****	**0.1676***	**−0.3677*****	**−0.3132****	**0.2894****
Stability	-	**0.3052***	0.1203	**0.3875*****	**0.3437****	**−0.3563*****
Balance	-	-	0.0659	**0.3404*****	**0.3281*****	**0.3693*****
Liking	-	-	-	−0.0099	−0.0542	−0.0533
	Medium complexity: *P1, M4, M5*: 					
	Stability	**Balance**	**Liking**	**APB**	**DCM**	**VDCM**
Dynamics	**−0.6289*****	**−0.4459*****	**0.2170***	**−0.2144***	**−0.2276*****	**0.4058*****
Stability	-	**0.6797*****	0.1295	**0.1957***	**0.2333*****	**−0.3218*****
Balance	-	-	0.1291	**0.2087***	**0.2473***	**−0.2715***
Liking	-	-	-	0.1535	0.0736	−0.0074
	Low complexity: *M1, M2, M3*: 					
	Stability	**Balance**	**Liking**	**APB**	**DCM**	**VDCM**
Dynamics	**−0.6869*****	**−0.5735*****	−0.1620	**−0.1890***	**−0.4598*****	**0.5948*****
Stability	-	**0.6723*****	0.0738	**0.2788***	**0.4253*****	**−0.4441*****
Balance	-	-	0.0229	**0.2183***	**0.3679*****	**−0.4274*****
Liking	-	-	-	0.1038	0.0277	0.0196

For artworks of high, low and medium complexity representative dynamic compositions of participants are shown. APB, DCM and VDCM were computed as described above. Significance is indicated as follows $p < .05$* $p < .001$**$p < .0001$***. The VDCM can take on negative and positive values. Note that for the VDCM, a negative correlation with stability means that arrangements with smaller values of VDCM were perceived as more stable than those with higher VDCMs. We did not include correlations with HDCM in Table 1 since these were always non-significant. (See also in the Appendix Figure A1 for average VDCM and HDCM values for each composition, Figure A2 for a graphical overview of the correlation values, and Table A2 for correlations across all stimuli, not separated by complexity level). APB = assessment of preference for balance; HDCM = horizontal deviation of the center of mass; VDCM = vertical deviation of the center of mass; DCM = deviation of the center of mass.

$F(1,836) = 2.35$, $p = .126$, and a significant interaction $F(1,836) = 647.74$, $p < .0001$. The interaction was the expected outcome here provided the composition tasks had been completed reasonably: We expected a classic disordinal pattern: In the dynamic task, ratings for dynamics should be high and for stability low. The opposite pattern should occur for stable compositions. Table A1, in the Appendix, shows numerical values for mean ratings of stability and dynamics for every composition.

Preference and Perception of Self-Arranged Compositions

Since we know that the relationships between APB and DCM scores and perceptual judgments varies depending on the number of elements present in an image (e.g., Fillinger & Hübner, 2020a) we performed our correlation analysis separately for the three groups of image complexity. Table 1 reveals that, indeed, the relationship between DCM, and APB scores and other perceptual measures tended to vary with the complexity of the artwork. For example, the positive correlation between dynamics and liking was strongest for medium complexity stimuli. In contrast, Fillinger and Hübner (2020a), found a strong negative correlation between movement and liking for stimuli of comparable complexity. Table 1 also shows that DCM and APB scores correlated negatively

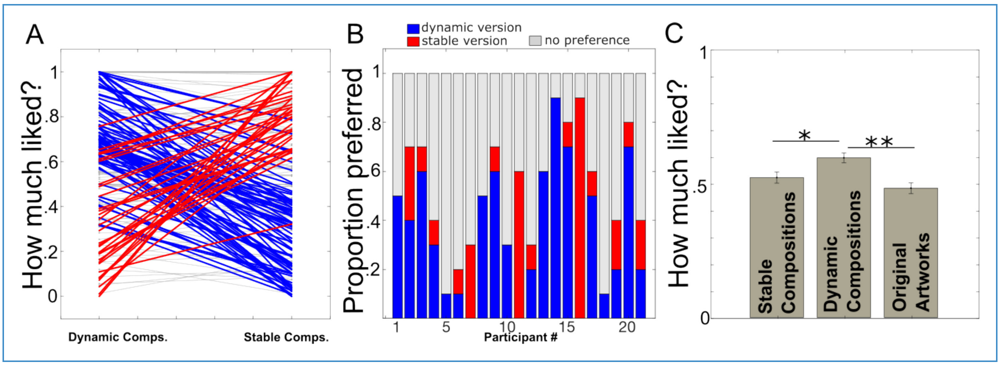


Figure 7. Individual preferences for dynamic and stable compositions. (A) Colored lines represent composition pairs where the dynamic (blue) composition was more liked by a participant than the corresponding stable (red) version. The resulting x-shape of the blue and red lines connecting the 10 composition pairs reflects the individual differences in liking of stable and dynamic versions. Gray lines represent composition pairs with no preference dynamic or stable versions. There are 210 (10 pairs×21 participants) dynamic stable composition pairs (lines) in this plot. The cut-off value for preference was a 0.2 difference between the participant's liking of a given dynamic and stable compositions. (See text for details.) (B) Proportion (out of 10) of preferences (liking) for dynamic and stable compositions for each participant, as well as the proportion of no preference (all three proportions adding up to one). Note, that "no preference" does not imply acting randomly. Instead, it suggests that stable and dynamic compositions were liked equally. Colors as in (A). (C) Overall, dynamic self-arranged compositions were liked better than stable compositions and original artworks (*$p < .05$, **$p < .001$). There was no difference in liking between stable compositions and original artworks.

with dynamics, and positively with stability and balance for stimuli of high- medium- and low-complexity, again opposite to what had been reported in previous work (although those correlations did not reach significance in Fillinger & Hübner, 2020a). Overall, the correlations of all ratings with VDCM tended to be stronger than with DCM. In general, and across all complexity levels, correlations between dynamics and VDCM and stability were positive, and negative, respectively, however was strongest for compositions of medium complexity. Interestingly, and in contrast to mentioned earlier work, neither DCM, nor VDCM, nor APB scores correlated with the amount of liking for any of the complexity levels. The overall somewhat weaker correlations of APB and DCM with ratings —compared to previous research—may be in part due to the substantial variability of APB and DCM indices in the self-arranged compositions across observers.

Some Participants Prefer Dynamic Compositions

Given the diversity of like rating data (Figure 6) we suspected that participants might differ in their preferences for composition "types." If true, we would expect that participants who prefer dynamic compositions, do not like stable ones and vice versa. To test this idea, we compared liking between stable and dynamic self-arranged compositions. Figure 7A shows all participants' liking scores for all dynamic and stable composition pairs (e.g., for $M1$, $M2$ etc). We quantify a preference for dynamic compositions to be a rating difference of at least 0.2 for a dynamic—stable composition pair (e.g., $M1$). If this difference passes the cutoff and is positive, the participant is labeled for this composition as having a "dynamic" preference, if this difference is negative and passes the cut-off, as having a stable preference for this composition. Figure 7A shows liking for dynamic and stable version of a composition, where the two liking scores of individual participants are connected

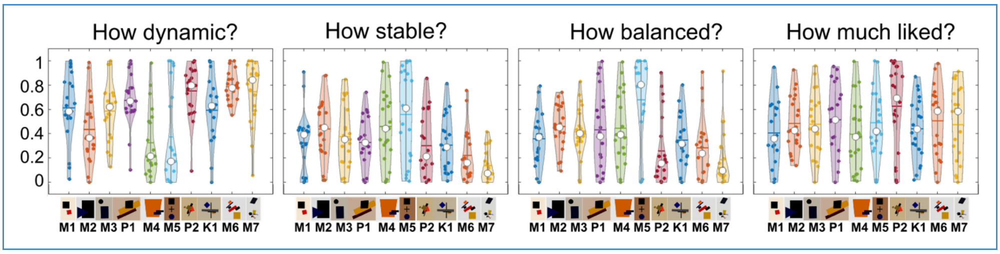


Figure 8. Rating data for the ten original suprematist compositions. As in Figure 6 colored dots correspond individual participants' data, white dots denote the median and solid horizontal lines the mean. Along the *x*-axis data for the individual artworks are represented as small images and by their shorthand. The *y*-axis indicates the rating strength for a given attribute/task (see Methods).

with a line. Dynamic preferences for a given composition are colored in blue (i.e., a blue line), and preferences for stable compositions are colored in red. No preference pairs are colored light gray, that is, those that did not pass the cutoff value. Figure 7A shows an "X" shaped pattern in this overall preference plot, which suggests that there may differences in how participants prefer stable and dynamic compositions, conversely, it could also mean that certain artworks are preferred as dynamic/stable version.

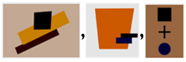
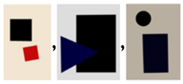
To tease these two possibilities apart we need to quantify how consistent participants are in their preferences across all ten compositions pairs. To this end, we computed the number of times a participant preferred stable over dynamic (red) or vice versa (blue) for the ten dynamic-stable composition comparisons and plotted these numbers as bars in Figure 7B. Blue bars denote a preference for dynamic compositions, red bars for stable ones, gray bars for no preference. As can be seen, several participants tended to have a preference for either the stable or dynamic version of a compositions, for example, 11 participants (no. 1, 3, 8, 9, 10, 11, 13, 14, 15, 17, 20) for dynamic and three participants (no. 7, 11, 16) for stable arrangements. Seven participants (no. 5, 6, 18, 19, 21) had no, or only a weak preference for one of the composition types.

Preference and Perception of the 10 Original Suprematist Artworks

We repeated the above set of analyses for ratings of the original artworks. Figure 8 shows that participants' ratings varied across artworks and tasks. A 10 (artworks)×4 (rating task: Dynamic, stable, balance, like) ANOVA supports this interpretation yielding a main effect of artworks $F(9,800) = 2.11$, $p = .026$, rating tasks $F(3,800) = 53.28$, $p < .001$, and an interaction $F(27,800) = 8.32$, $p < .001$. The interaction underlines that the patterns of ratings vary as a function of artwork: Some artworks are rated as more dynamic and less stable/balanced (e.g., *M*6), whereas others are rated as less dynamic and more stable (e.g., *M*5). While individual ratings in the rating tasks (dynamic, stable, balance) seem to be somewhat clustered around values, "likings" appear to be, again, broadly distributed, which suggests that there was not much agreement between participants on whether they liked an artwork or not. Overall, original artworks were liked less than self-arranged compositions ($M_{original}$=0.481, $M_{dynamic} = 0.595$, $M_{stable} = 0.5211$). A 3 (original, dynamic, stable)×10 (artworks) ANOVA yielded a significant main effect of composition $F(2,600) = 8.647$, $p = .0002$, for post hoc comparison results see Figure 7C), no main effect of artwork and no interaction.

Table 2 shows that the relationship between ratings and geometric indices varies substantially with the complexity of the artwork, and more so than in Table 1. For example, the correlation between stability and DCM is negative for high complexity artworks, is positive for those of medium complexity and insignificant for artworks of low complexity. Only two relationships

Table 2. Correlations between APB, DCM scores and ratings of original suprematist artworks sorted by complexity.

	High Complexity: *P2, K1, M6, M7*: 					
	Stability	Balance	Liking	APB	DCM	VDCM
Dynamics	**−0.3506***	−0.1007	−0.0078	−0.0699	0.1649	−0.0265
Stability	-	**0.3052***	**0.4607*****	**0.2707***	**−0.2625***	−0.1476
Balance	-	-	0.0976	0.1954	−0.1015	−0.1609
Liking	-	-	-	0.0686	−0.1028	−0.0177
	Medium complexity: *P1, M4, M5*: 					
	Stability	**Balance**	**Liking**	**APB**	**DCM**	**VDCM**
Dynamics	**−0.4081****	**−0.3928****	0.2055	**−0.4576****	**−0.4676****	**−0.4240****
Stability	-	**0.5454*****	**0.333****	**0.2855***	**0.2616***	**0.3025***
Balance	-	-	0.1522	0.2074	0.1562	**0.2626***
Liking	-	-	-	−0.1052	−0.1291	−0.0702
	Low complexity: *M1, M2, M3*: 					
	Stability	**Balance**	**Liking**	**APB**	**DCM**	**VDCM**
Dynamics	**−0.4522****	−0.1172	−0.1416	**0.2907***	**0.2538***	0.1253
Stability	-	**0.3695****	0.1013	−0.1820	−0.1807	−0.1270
Balance	-	-	0.1761	−0.1966	−0.2048	−0.1584
Liking	-	-	-	−0.1113	−0.1380	−0.1388

Shown are the correlations as in Table 2, here for original Suprematist artworks of high, low- and medium-complexity. APB, DCM and VDCM were computed as described above. Significance is indicated as follows $p < .05$* $p < .001$** $p < .0001$***. Overall, the relationship between DCM and APB indices and the ratings varies substantially with the complexity of the artwork. Only for stimuli of medium complexity balance and dynamics correlated negatively and significantly. DCM indices correlated negatively with stability for artworks of high complexity and positively with those of medium complexity. They also correlated negatively with dynamics for medium complexity artworks and positively for low complexity artworks. APB indices correlated positively with stability for high and medium complexity artworks but not for low complexity artworks. APB indices also correlated negatively with dynamics for stimuli of medium complexity and correlated positively with dynamics for low complexity stimuli. Liking correlated positively with stability only for high- and medium-complexity artworks. In the Appendix, Figure A3 provides a graphical overview how correlations change across complexity levels. Also see Table A4 in the Appendix for correlations across all stimuli (not separated by complexity level). APB = assessment of preference for balance; VDCM = vertical deviation of the center of mass; DCM = deviation of the center of mass.

remained unchanged across complexity levels: The negative correlation between dynamics and stability, and the positive correlation between stability and balance. Also unchanged across complexity level was, that liking and balance did not correlate with each other and neither with the geometric indices. This is further illustrated in Figure A4 in the Appendix, in which ratings are color-codes by complexity group and plotted as a function of DCM and APB. The lack of structure in the data is apparent and suggests that the geometric indices only partly relate to overall ratings and liking, and that large individual differences in perceptual assessments and liking exist.

Individuals Have Different Compositional Preferences

Given the diversity of liking data for the original Suprematist compositions we suspected that participants might again differ in their compositional preference. If true, we would see that participants who prefer dynamic compositions also do not like stable ones and vice versa. To test this idea we picked, based on the average ratings (Figure 8), the most and least dynamic (*M6*: 0.788, *M4*: 0.325),

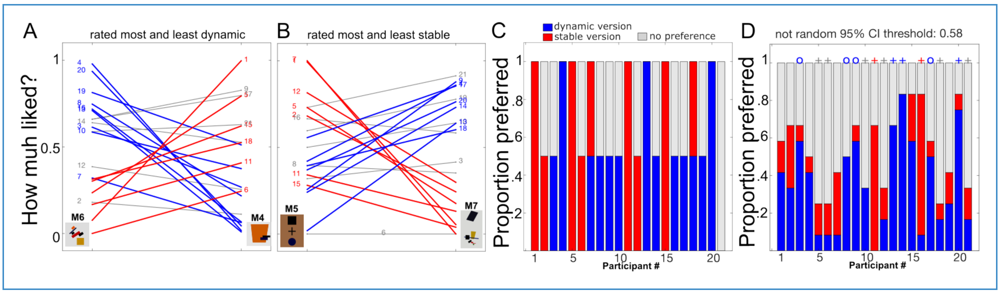


Figure 9. Individual preferences for dynamic and stable artworks. (A) Participants with a preference for dynamic compositions are represented with a blue line, those for stable compositions with a red one. Numbers refer to specific participants. Participants without a strong preference are represented with gray lines. The cutoff value for preference was a 0.2 difference between the participant's liking of the most (*M6*) and least dynamic (*M4*) Suprematist artwork (See text for details.) (B) Preferences (liking) for the most (*M5*) and least stable (*M7*) artworks are shown. Colors and symbols as in (A). (C) Bar length corresponds to consistency score in preference. Red and blue colors denote preferences for stable and dynamic compositions, respectively. Gray means no preference. Each bar stands for a single participant identified by a number. Remember that for this analysis we consider like ratings for 2 "pairs" of stimuli. A blue bar of length 1 means that the participant consistently preferred the more dynamic and less stable compositions in (A) and (B), respectively (i.e., 2/2). A red bar of length 1 means that the participant consistently preferred the more stable and less dynamic compositions in (A) and (B), respectively. Bars of length 0.5 mean that the participant preferred the more dynamic (blue) or stable (red) composition in one set (e.g., A or B) and showed no preference (gray portion of the bar) in the other. Bars that are half blue and half red indicate that this participant switched preference. (D) Formal assessment of randomness of preference patterns across all stimulus pairs: 10 dynamic-stable self-composition pairs + 2 original stable-dynamic pairs. Any gray, red, or blue portion of the bar exceeding the threshold of 0.58 suggests that this person' preference was significantly different from random (marked with a "+"). Participants' whose proportion of preference for one composition type (stable or dynamic) was significantly (about 6 times) larger than that of the other composition type (stable or dynamic), are marked with a "o." Participants with no markings above the bar indicated their preferences not different than a random observer. See text for details.

and the most and least stable artworks (*M5*: 0.559, *M7*: 0.137, see also Table A3 in the Appendix for means of all artworks), and compared liking patterns for those. In Figure 9, liking patterns are plotted for all observers for the most and least dynamic rated (panel A) and the most and least stable rated artworks (panel B). We use again the same color-code as in Figure 7 and plot the data of participants preferring stable compositions in red and data of those preferring dynamics ones in blue. As before we use the same cutoff difference of 0.2 and assume that a "stable preferring" participant should prefer stable compositions by at least 0.2 over dynamic compositions and "dynamics preferring" participants should prefer dynamic compositions over stable ones. Inspecting Figures 9A and B we see, again, differences in preferences between participants as evident by the "x" pattern in the plot, for the most-least dynamic, and most-least stable rated stimulus pairs, respectively and Figure 9C reports the proportion dynamic or stable preferences for each participant. Seven participants stayed perfectly consistent (bars with height 1), 11 had a preference only for one of the stimulus pairs and were labeled "indifferent" in the other (bars half colored, half gray). One participant (21) had no preference (gray bar with height 1) and two participants (7 and 18) were inconsistent in their preference (bars half red and half blue), for these stimulus pairs.

But are these preference patterns for self-arranged and original compositions just random or systematic? To formally test this, we simulated an observer whos' liking setting would vary according to a uniform random variable of the interval [0 1], that is, completely independent of what the actual stimulus is. Similar to our experiments, we obtained 12 paired simulated like values from our random observer (in the experiment: 10 pairs for self-arranged compositions, 2 pairs from the original compositions). Then, as in the analysis of the real data, we computed the differences between simulated like values of a given pair, compared it to a cutoff (0.2), noted the preference (option 1 or 2 in a given pair) and counted the number of preferences for option 1, option 2, and number of no preferences. We repeated this simulation 100,000 times and recorded for each the three preference counts (1, 2, no). This will yield three distributions of preference counts as they occur for a random-acting observer (See Appendix Figure A6). These are approximately normally distributed, and the means tell us, how often, on average, a random observer prefers option 1, option 2, or has no preference. We can also determine extreme values, for example, those that are extremely unlikely to occur for a random observer—such as those in the 5% tail of the distribution. This is simply achieved by sorting the counts for each preference option (1,2, no) according to size and then find the value of the 95th percentile. If a given measurement exceeds that value, it means we can be 95% confident that it did not occur due to chance. In our simulation that value was 0.583. In Figure 9D, where preference proportions for each observer are shown, we labeled the cases with such clear preferences with an "+." Those with a gray "+" have no preference (5,6,10,12,18,21), that is, they like dynamic and stable compositions equally (much or little), those with a red or blue "+" have a clear preference for either stable (11,16) or dynamic (13,14,20) compositions, respectively.

Next, we also determined a threshold for the ratio of preferences. The goal was to understand, that even if preferences for dynamic and stable did not exceed the above cutoff, whether the proportion of stable-dynamic preferences exceeded the threshold for randomness. To this end we computed the ratio of the larger proportion preference to the smaller proportion preference. Then, as above, we sort the resulting distribution of ratios and determine the value corresponding to 95th percentile. This value was 5.994, meaning that the preference of one option over the other had to be six times as much to be significantly different from random. The preference for dynamic compositions of four observers (3,8,9,17) exceeded this value are marked with a blue "o" in Figure 9D. This leaves six observers (1,2,4,7,15,19), for which we cannot reject the possibility that their indicated preferences followed a random pattern.

Discussion

Summary

We investigated with two interactive tasks and artworks of the Suprematist movement whether the relationships between perceived stability, dynamics, and liking in self-arranged compositions vary across complexity levels for art-naïve participants. First, we asked participants to create compositions that either appeared stable or dynamic to them. The materials used for the tasks consisted of simple geometric colored shape elements and their background "canvases" taken from ten original Suprematists' artworks. In subsequent rating experiments containing both original and self-arranged compositions, participants rated how dynamic, stable, and balanced they perceived the compositions and how much they liked them. We found that participants are able to create dynamic and stable compositions and are quite consistent with respect to their individual preferences and rating behaviors. Participants had a sense for the composition types and rated their stable compositions as predominantly stable, their dynamic compositions as predominantly dynamic and liked their self-arranged compositions more than the original Suprematist artworks. For self-arranged compositions the number of shape elements present in the composition only weakly modulated correlations between the ratings of dynamics, stability, and balancedness, and liking and geometrical measures for balance. For all three levels of complexity, APB and DCM indices correlated positively with perceived balance and stability and negatively with perceived dynamics, a pattern different to that reported in the literature (Fillinger & Hübner, 2020a). For the original Suprematist artworks we found that correlations between the three ratings, liking and geometrical indices changed substantially across complexity levels, even changing sign, and overall, no relationship between DCM and APB indices and perceived balance, stability and dynamics was found. One striking result was the lack of consensus between participants how much they liked their self-arranged and the original compositions. We hypothesized that this spread in liking scores might be due to individual differences in preference for stable or dynamic compositions. An analysis of rating patterns revealed individual preferences for several participants with a large group preferring dynamic compositions.

Participants Have a Sense of Overall Compositional Quality and are Able to Create Dynamic and Stable Compositions

With an interactive approach, we wanted to investigate the concepts of stability and dynamics in compositions. For our study, we selected 10 artworks from the Suprematist movement and tested 21 art-naïve participants What makes this art so fascinating, is its strong reduction to just few shape elements colored with a very limited range of hues. This extreme reduction of elements and colors allows to investigate a lay person's sensitivity for different compositional aspects such as effects of shape arrangements within a pictorial field. Minimalistic changes of the position or orientation of a single shape or few shapes within the pictorial field can give rise to very different impressions of the composition, to the perceived balance and whether the composition evokes a more static or dynamic impression (e.g., Figures 1 and 4). In order to make such an open-ended, potentially difficult task feasible to art-naïve participants—without having to create actual paintings—we provided them with the basic elements of a composition, that is, different numbers of colored shape elements that had to be arranged on a background of a certain size and color—all taken from a corresponding original Suprematist artwork. Participants enjoyed the composition tasks, tested the effects of different shape arrangements and potentially emulating the process of "creative pains" artists have to go through to find the "perfect" positions and orientations for elements of the composition. However, in our experiment the tasks for our participants were much simpler: Not perfection was the goal but a stable or a dynamic composition for the elements of ten artworks of different complexity.

Inspecting participants' arrangements in Figure 4, and also ratings of these compositions (Figure 5), shows, that most participants performed these tasks reasonably well. *Dynamic* compositions were rated on average to be more dynamic than *stable* ones and *stable* compositions were rated as more stable than *dynamic* ones. However, ratings for stable compositions tended to be closer to the extreme point of the rating scale (i.e., closer to 1) than ratings for dynamic compositions. This, and also the increased variance in ratings for dynamic arrangements might suggest that creating stable compositions might have been an easier task for our participants. In other words, the concept and/or sense of stability might be more straight-forward and more similar between participants than the concept and/or sense of dynamics. Consistent with this, several participants reported in the exit survey that they indeed perceived the stable task to be "easy" and "intuitive."

In Figure 4, we also see that stability was interpreted as some kind of gravitational stability. Participants frequently "built up" their shape arrangements from the bottom of the backgrounds, culminating in "stacked" and often symmetric compositions. This is a quite reasonable approach, as many artists, for example Rothko, Klee, Hoffmann created very stable compositions that indicate no dynamics or movements by arranging the main orientation of shaped contours or color fields parallel to the lower border of the canvas. Symmetry is a common and important way to produce balance in nature, and symmetric compositions are perceived as balanced (Eisenman & Gellens, 1968; Fillinger & Hübner, 2020a; Hübner & Thömmes, 2019).

Not only position and orientation but also the color of the shape elements may modulate how dynamic or stable a composition is perceived. For example, McManus et al. (1985), using abstract displays of one or two colored squares (red, green, or blue), found that there were also significant interactions of position with color, with red affecting balance the most and while blue the least. Locher et al. (2005) found such effects of color when comparing perceived balance in Piet Mondrian's original abstract paintings versus variants with interchanged colors. Observers judged the originals to be balanced near the center and to be more balanced than the color variants. Based on judgments of "balance center," the perceived weight of a shape varied as a function of its size and color, with red being perceived as heaviest, blue intermediate, and yellow lightest (Locher et al., 2005). While we did not vary color systematically nor analyzed this formally, inspecting the stable self-arranged compositions we see that the size of a shape element more strongly affected the choice for its position in the composition than color. In most compositions larger elements tended to be placed near the lower edge of the canvas, creating some kind of basement on which smaller elements were stacked. Consequently, we find "lighter" yellow/orange at base positions (*M*4, *M*6) and red elements at the top (*M*1)—which visually should evoke some kind of imbalance.

Stacked, symmetric arrangements, as we see them for the stable composition task also led to asymmetries in the vertical direction, and consequently, to larger VDCM (Figure A1), DCM, and APB values. For dynamic arrangements, participants rotated shape elements and created wider, more uneven compositions. These arrangements were rated as much less balanced than stable ones, yet, because of the spread of elements APB, and DCM values tended to be significantly smaller for dynamic compositions than for stable ones. This suggests that, in a numerical sense, dynamic compositions were more balanced than stable ones. Yet, the more stable a composition appeared, the more balanced it was rated by the participants (Tables 1 and 2, Appendix Table A1). This result contradicts the idea of an objective geometric measure of balance (e.g., Wilson & Chatterjee (2005) and Hübner & Fillinger (2016)) and shows instead that APB and DCM indices, which are mainly measures of pixel distributions in relation to the cardinal axes or the center of the image (and thus are related to the concept of "mechanical balance"), do not fully capture all concepts of perceptual balance, such as the sense of stability (also see Fillinger & Hübner, 2020a, p. 2020b; Hübner & Fillinger, 2019). The negative correlation of perceived stability with the VDCM (Table 1), and the strong positive correlation between perceived balance and perceived stability support this interpretation.

It is interesting to note, that for original Suprematists' artworks the relationships between the geometric indices and ratings change even more across complexity levels: They change strengths and even in sign and/or become non-significant. For example, the correlation of APB and indices with perceived stability is negative for high-, positive for medium-, and absent for low-complexity compositions. Modulations of the relationship between perceived compositional qualities such as balance, stability and dynamics and image-computed indices of balance have also been reported in earlier work (e.g., Fillinger & Hübner, 2020a).

This suggests that the geometrical indices are not particularly sensitive to subtle balance differences as they might exist in the original artworks, but instead capture more coarse aspects of compositions. In our study the participant's compositions were created with two very distinct goals: Achieving a stable vs. achieving a dynamic composition with the same shape elements, yet perhaps our range of stimuli (in terms of number of elements and colors) was much more varied than the set used by Fillinger and Hübner (2020a) which might explain discrepancies between our findings and theirs.

Also, the relationship between perceived dynamics, stability and balance varied with the number of shape elements used in a composition. For example, we found that the correlation between stability and balance was about half as strong for high- and low-complexity artworks than that for those of medium complexity. This suggests that participants might have used somewhat different criteria to assess balance for compositions with many or very few shape elements, or the task was too difficult for them, for example, participants could not achieve a stable composition that also appeared balanced to them.

Aesthetics Preference and the Perception of a Composition

Self-arranged dynamic compositions were overall more liked by the participants compared to their stable compositions, and the correlations between liking and dynamics were positive for compositions of medium and high complexity. These were liked more, the more dynamic their compositions were perceived. These relationships between liking and dynamics, and liking and stability were absent for compositions with just one or two elements. In contrast to this, Fillinger and Hübner (2020a) found that for compositions with few elements, the more balanced and stable a composition was perceived, the more it was liked. For more complex artworks, they found, like us, a positive correlation between liking and movement, however, their correlation was much stronger (.549 vs .1676 (here)). Fillinger and Hübner (2020a) suggested that participants evaluated complex compositions that implied motion differently than those that didn't, for example, instability associated with gravitation was not liked, whereas instability associated with movement was liked. Our results confirm, that this is especially the case if more shape elements are present in a composition. Our participants differed in their preference for dynamic over stable arrangements. Strikingly, not only do most of the participants have a certain preference (Figures 7 and 9), but these preferences are rather stable.

For the original Suprematist artworks, however we do not find a relationship between liking and dynamics for any of the three levels of complexity, despite having the same number of shape elements, colors, and backgrounds and even though quite a few of the compositions were rated as dynamic (Figure 8). Instead, participants liked a composition more, the more stable they perceived it (Table 2)—especially for artworks of high and medium complexity. The fact that we don't find strong correlations between any of the perceptual measures and like is not surprising given the large spread of liking data for dynamic, stable compositions and original compositions (Figures 6 and 8, right most panels). One may conclude that this art is not exactly every one's cup of tea. However, we suspected and tested the idea that the spread in liking could mean that participants might consist of subgroups that differ in their aesthetic preferences.

Where might such preferences come from? Although we tested only non-experts different preferences may be explained by some kind of training. For example, while many studies found that people in general prefer symmetrical and complex stimuli more than asymmetrical and simple ones (Gartus & Leder, 2013; Höfel & Jacobsen, 2007; Jacobsen & Höfel, 2002; Tinio & Leder, 2009), Leder et al. (2019) compared the beauty ratings of art experts and non-experts with stimuli of Jacobsen and Höfel (2002) that varied in symmetry and complexity and found that experts rather preferred simple and asymmetric stimuli, thus expertise and training in art can change perceptual preferences and aesthetic appreciation. Participants in our study were not art experts and general patterns of complex over simple compositions were not apparent, therefore other factors than training should account for why individuals differ in their aesthetical preferences. It has been shown for example, that aesthetic preferences might be linked to the specific art and topic, as well as individual life-times experiences and even differences in neural make-up (Eysenck, 1940; McManus, 1980; McManus et al., 1981; Palmer et al., 2013; Van Geert & Wagemans, 2020; Vessel et al., 2018; Williams et al., 2018). To uncover these links remains the subject of future study.

Conclusions

Using works of the Suprematist artist and an interactive experimental approach we investigated what art-naïve-people understand when creating a stable and a dynamic composition. We found that non-experts can do these tasks and we discovered consistent individual differences in the preferences for stable and dynamic compositions that were reflected in the ratings of self-arranged compositions and the original artworks. The existence of such individual differences should be factored in when assessing “liking” in studies of empirical aesthetics.

Acknowledgements

The authors thank Vanessa Braner for her work and enthusiasm in selecting and measuring the original artworks and for fruitful and inspiring discussions. We also thank Victoria Neuwirt for her help, diligence, and responsibility in data collection, and our observers for participating in the experiments and playing with the shapes until a stable or dynamic composition was achieved. We thank Denis Pelli for helpful discussions, as well as Ian C. McManus and an anonymous reviewer for constructive comments and helpful suggestions. We also want to thank the artists of the Suprematist art movement for their works in the attempt to achieve “the supremacy of pure feeling.” This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; Project No. 222641018–SFB/TRR 135 Projects A1 and B8), and by the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project “The Adaptive Mind”).

ORCID iDs

Doris I. Braun https://orcid.org/0000-0002-6626-0572
Katja Doerschner https://orcid.org/0000-0002-8364-800X

Author Contribution(s)

Doris I. Braun: Conceptualization; Data curation; Formal analysis; Formal analysis; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Mara Hofmann: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing – original draft.

Katja Doerschner: Conceptualization; Data curation; Formal analysis; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; Project No. 222641018–SFB/TRR 135 Projects A1 and B8), and by the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project “The Adaptive Mind”).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abeln, J., Fresz, L., Amirshahi, S. A., McManus, I. C., Koch, M., Kreysa, H., & Redies, C. (2016). Preference for well-balanced saliency in details cropped from photographs. *Frontiers in Human Neuroscience*, *9*, 704. https://doi.org/10.3389/fnhum.2015.00704

Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye* (pp. xii, 408). University of California Press.

Arnheim, R. (2009). *The Power of the Center,A Study of Composition in the Visual Arts (20th Anniversary Edition)*. University of California Press. https://www.ucpress.edu/books/the-power-of-the-center/paper

Belke, B., Leder, H., & Augustin, M. D. (2006). Mastering style—effects of explicit style-related information, art knowledge and affective state on appreciation of abstract paintings. *Psychology Science*, *48*(2), 115–134.

Braun, D. I., & Doerschner, K. (2019). Kandinsky or me? How free is the eye of the beholder in abstract art? *I-Perception*, *10*(5), 2041669519867973. https://doi.org/10.1177/2041669519867973

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, *18*(7), 370–375. https://doi.org/10.1016/j.tics.2014.03.003

Corradi, G., Chuquichambi, E. G., Barrada, J. R., Clemente, A., & Nadal, M. (2019). A new conception of visual aesthetic sensitivity. *British Journal of Psychology*, *111*(4), 630–658. https://doi.org/10.1111/bjop.12427

Eisenman, R., & Gellens, H. K. (1968). Preferences for complexity-simplicity and symmetry-asymmetry. *Perceptual and Motor Skills*, *26*(3, PT. 1), 888–890. https://doi.org/10.2466/pms.1968.26.3.888

Eysenck, H. J. (1940). *The 'general factor' in aesthetic judgements*.

Fechner, G. T. (1871). *Zur experimentalen aesthetik [electronic resource] /*. S. Hirzel,. http://archive.org/details/ldpd_7460548_000

Fechner, G. T. (1876). *Vorschule der Aesthetik*. Breitkopf & Härtel.

Fillinger, M. G., & Hübner, R. (2020a). On the relation between perceived stability and aesthetic appreciation. *Acta Psychologica*, *208*, 103082. https://doi.org/10.1016/j.actpsy.2020.103082

Fillinger, M. G., & Hübner, R. (2020b). Relations between balance, prototypicality, and aesthetic appreciation for Japanese calligraphy. *Empirical Studies of the Arts*, *38*(2), 172–190. https://doi.org/10.1177/0276237418805656

Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of broken symmetries. *I-Perception*, *4*(5), 361–364. https://doi.org/10.1068/i0588sas

Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *I-Perception*, *2*(6), 508–527. https://doi.org/10.1068/i0472

Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test: I. Construction and psychometric properties. *Perceptual and Motor Skills*, *49*(3), 795–802. https://doi.org/10.2466/pms.1979.49.3.795

Höfel, L., & Jacobsen, T. (2007). Electrophysiological indices of processing symmetry and aesthetics: A result of judgment categorization or judgment report? *Journal of Psychophysiology*, *21*(1), 9–21. https://doi.org/10.1027/0269-8803.21.1.9

Hübner, R. (2025). Preference for symmetry, balance, or proximity in picture aesthetics depends on the method of evaluation. *I-Perception*, *16*(6), 20416695251381548. https://doi.org/10.1177/20416695251381548

Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, *7*, 1–15. https://doi.org/10.3389/fpsyg.2016.00335

Hübner, R., & Fillinger, M. G. (2019). Perceptual balance, stability, and aesthetic appreciation: Their relations depend on the picture type. *I-Perception*, *10*(3), 204166951985604. https://doi.org/10.1177/2041669519856040

Hübner, R., & Thömmes, K. (2019). Symmetry and balance as factors of aesthetic appreciation: Ethel puffer's (1903) "studies in symmetry" revised. *Symmetry*, *11*(12), 1468. https://doi.org/10.3390/sym11121468

Ishihara, S. (2018). *Ishihara's test for colour deficiency: 24 plates edition*.

Jacobsen, T., & Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills*, *95*(3 Pt 1), 755–766. https://doi.org/10.2466/pms.2002.95.3.755

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Ein Beitrag zur Analyse der malerischen Elemente* (2. Aufl., unverändert). https://doi.org/10.11588/diglit.29206

Koenderink, J., Van Doorn, A., & Gegenfurtner, K. (2018). Color weight photometry. *Vision Research*, *151*, 88–98. https://doi.org/10.1016/j.visres.2017.06.006

Leder, H. (2002). *Explorationen in der Bildästhetik: Vertrautheit, künstlerischer Stil und der Einfluss von Wissen als Determinanten von Präferenzen bei der Kunstbetrachtung*. Pabst Science Publ. https://scholar.google.com/scholar?cluster=9355037240212649507&hl=en&oi=scholarr

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology (London, England: 1953)*, *95*(Pt 4), 489–508. https://doi.org/10.1348/0007126042369811

Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 2–10. https://doi.org/10.1037/a0026396

Leder, H., Tinio, P. P. L., Brieber, D., Kröner, T., Jacobsen, T., & Rosenberg, R. (2019). Symmetry is not a universal law of beauty. *Empirical Studies of the Arts*, *37*(1), 104–114. https://doi.org/10.1177/0276237418777941

Locher, P. (2003). Experimental techniques for investigating the contribution of pictorial balance to the creation and perception of visual displays. *Acta Psychologica*, *21*(2), 147–164. https://doi.org/10.2190/0QDJ-H0LP-PBT0-D4DY

Locher, P., Cornelis, E., Wagemans, J., & Stappers, P. J. (2001). Artists' use of compositional balance for creating visual displays. *Empirical Studies of the Arts*, *19*(2), 213–227. https://doi.org/10.2190/EKMD-YMN5-NJUG-34BK

Locher, P., Overbeeke, K., & Stappers, P. J. (2005). Spatial balance of color triads in the abstract art of piet mondrian. *Perception*, *34*(2), 169–189. https://doi.org/10.1068/p5033

Locher, P. J., Stappers, P. J., & Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychologica*, *99*(2), 141–161. https://doi.org/10.1016/S0001-6918(98)00008-0

Marković, S. (2011). Perceptual, semantic and affective dimensions of experience of abstract and representational paintings. *Psihologija*, *44*(3), 191–210. https://doi.org/10.2298/PSI1103191M

McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, *71*(4), 505–524. https://doi.org/10.1111/j.2044-8295.1980.tb01763.x

McManus, I. C. (2005). Symmetry and asymmetry in aesthetics and the arts. *European Review*, *13*(S2), 157–180. https://doi.org/10.1017/S1062798705000736

McManus, I. C., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, *76*(3), 311–324. https://doi.org/10.1111/j.2044-8295.1985.tb01955.x

McManus, I. C., & Gesiak, P. (2020). *I C McManus and Paul Gesiak: Experimenting with Mondrian: Comparing the Method of Production with the Method of Choice*. *Poster—Design Computing and Cognition 2014, University College London* (No. n6cf7_v1). PsyArXiv. https://doi.org/10.31234/osf.io/n6cf7

McManus, I. C., Jones, A. L., & Cottrell, J. (1981). The aesthetics of colour. *Perception*, *10*(6), 651–666. https://doi.org/10.1068/p100651

McManus, I. C., Stöver, K., & Kim, D. (2011). Arnheim's gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *I-Perception*, *2*(6), 615–647. https://doi.org/10.1068/i0445aap

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, *28*(2), 173–191. https://doi.org/10.2190/EM.28.2.d

Nath, S., Brändle, F., Schulz, E., Dayan, P., & Brielmann, A. (2023). *Relating Objective Complexity, Subjective Complexity and Beauty*. https://doi.org/10.31234/osf.io/nuep7

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, *64*(1), 77–107. https://doi.org/10.1146/annurev-psych-120710-100504

Pearce, M. T., Zaidel, D. W., Vartanian, O., Skov, M., Leder, H., Chatterjee, A., & Nadal, M. (2016). Neuroaesthetics: The cognitive neuroscience of aesthetic experience. *Perspectives on Psychological Science*, *11*(2), 265–279. https://doi.org/10.1177/1745691615621274

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Puffer, E. D. (1903). Studies in symmetry. *The Psychological Review: Monograph Supplement*, *4*, 467–539.
Ross, D. W. (with Robarts—University of Toronto) (1907). *A theory of pure design; harmony, balance, rhythm*. Boston, Houghton, Mifflin. http://archive.org/details/theoryofpuredesi00rossuoft
Sarami, A., Afhami, R., & Wagemans, J. (2021). Perceptual organisation affects perception and appreciation of abstract art: A case study with black square and red square by Kazimir Malevich. *Art & Perception*, *9*(4), 353–397. https://doi.org/10.1163/22134913-bja10028
Specker, E., Douda, M., & Leder, H. (2024). How do we understand artworks? Exploring the role of artwork inherent features in art processing. *Empirical Studies of the Arts*, *42*(2), 469–497. https://doi.org/10.1177/02762374231201074
Specker, E., Forster, M., Brinkmann, H., Boddy, J., Immelmann, B., Goller, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). Warm, lively, rough? Assessing agreement on aesthetic effects of artworks. *PLOS ONE*, *15*(5), e0232083. https://doi.org/10.1371/journal.pone.0232083
Specker, E., Tinio, P. P., & Van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(3), 265–275. https://doi.org/10.1037/aca0000107
Thömmes, K., & Hübner, R. (2018). Instagram likes for architectural photos can be predicted by quantitative balance measures and curvature. *Frontiers in Psychology*, *9*, 1–17. https://doi.org/10.3389/fpsyg.2018.01050
Tinio, P. P. L., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, *130*(3), 241–250. https://doi.org/10.1016/j.actpsy.2009.01.001
Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 135–154. https://doi.org/10.1037/aca0000224
Vessel, E. A., Maurer, N., Denker, A. H., & Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, *179*, 121–131. https://doi.org/10.1016/j.cognition.2018.06.009
Westphal-Fitch, G. (2019). Revisiting Fechner's methods. In M. Nadal & O. Vartanian (Eds.), *The Oxford handbook of empirical aesthetics* (1st ed., pp. 83–98). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.9
Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2007–2022. https://doi.org/10.1098/rstb.2012.0098
Westphal-Fitch, G., Oh, J., & Fitch, W. T. (2013). Studying aesthetics with the method of production: Effects of context and local symmetry. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 13–26. https://doi.org/10.1037/a0031795
Williams, P. G., Johnson, K. T., Curtis, B. J., King, J. B., & Anderson, J. S. (2018). Individual differences in aesthetic engagement are reflected in resting-state fMRI connectivity: Implications for stress resilience. *NeuroImage*, *179*, 156–165. https://doi.org/10.1016/j.neuroimage.2018.06.042
Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, *23*(2), 165–180. https://doi.org/10.2190/B1LR-MVF3-F36X-XR64

Appendix

Information A1. The Suprematist Art Movement. The art movement of Suprematism developed during the difficult social and political time period of the October Revolution, World War 1 and the outbreak of the Russian Civil War. In the early twentieth-century the Ukrainian artist and art theorist Kazimir Severinovich Malevich (Malevych) founded together with a group of mainly Russian artists including Liubov Popova, Lazar Khidekel, Aleksandra Ekster, Olga Razanova the art movement "Suprematism." The aim of this new non-objective art movement was "the supremacy of pure

Table A1. Mean ratings for participants' self-arranged compositions.

Dynamic	M1	M2	M3	P1	M4	M5	P2	K1	M6	M7
										
$M_{dynamic}$	0.623	0.708	0.739	0.578	0.515	0.629	0.700	0.648	0.734	0.714
$SD_{dynamic}$	0.272	0.208	0.185	0.378	0.317	0.308	0.269	0.324	0.287	0.298
$M_{stability}$	0.434	0.304	0.339	0.323	0.519	0.421	0.398	0.317	0.339	0.483
$SD_{stability}$	0.343	0.266	0.256	0.310	0.308	0.363	0.303	0.333	0.324	0.342
$M_{balance}$	0.466	0.321	0.374	0.424	0.464	0.414	0.397	0.350	0.440	0.415
$SD_{balance}$	0.330	0.338	0.310	0.335	0.230	0.328	0.301	0.327	0.362	0.310
M_{liking}	0.633	0.502	0.589	0.652	0.620	0.567	0.644	0.475	0.574	0.690
SD_{liking}	0.254	0.301	0.234	0.258	0.292	0.307	0.280	0.222	0.272	0.168
APB	44.85	19.62	39.78	29.34	21.52	43.74	38.37	39.99	36.24	42.09
DCM	28.31	13.88	24.30	21.11	12.67	26.97	29.56	33.11	26.72	35.68
Stable	***M1***	***M2***	***M3***	***P1***	***M4***	***M5***	***P2***	***K1***	***M6***	***M7***
										
$M_{dynamic}$	0.155	0.068	0.203	0.155	0.128	0.137	0.208	0.171	0.135	0.241
$SD_{dynamic}$	0.294	0.097	0.287	0.254	0.241	0.235	0.189	0.257	0.237	0.266
$M_{stability}$	0.914	0.853	0.871	0.852	0.887	0.831	0.782	0.775	0.849	0.747
$SD_{stability}$	0.134	0.202	0.162	0.191	0.188	0.199	0.242	0.278	0.268	0.291
$M_{balance}$	0.838	0.860	0.828	0.801	0.792	0.822	0.749	0.749	0.779	0.711
$SD_{balance}$	0.251	0.228	0.232	0.287	0.332	0.235	0.304	0.310	0.323	0.304
M_{liking}	0.514	0.482	0.464	0.475	0.582	0.600	0.526	0.456	0.542	0.571
SD_{liking}	0.334	0.291	0.255	0.338	0.324	0.307	0.258	0.241	0.350	0.300
APB	51.97	24.54	44.63	35.62	24.36	53.46	48.40	49.71	50.93	50.93
DCM	46.97	26.05	42.00	35.63	19.73	46.14	55.35	54.74	43.47	52.81

Shown are mean ratings and corresponding standard deviations for the *DYNAMIC* and *STABLE* compositions with shape elements of ten suprematist originals. Overall, compositions created with instruction to produce a dynamic arrangement were rated as more dynamic and less stable, compositions created with the instruction to produce a stable arrangement were rated in the opposite way. Dynamic compositions of low and high complexity tended to be rated slightly more dynamic (range: 0.623–0.739) than those of medium complexity (range: 0.515–0.629). APB and DCM are average values for DYNAMIC and STABLE self-arranged compositions. For APB and DCM, lower numbers indicate higher balance (possible range: 0%–100%). Stimuli are represented by the shorthand of the corresponding artwork and a representative participant composition. DCM = deviation of the center of mass; APB = assessment of preference for balance.

Table A2. Correlations between preference for balance (APB) and DCM and ratings of participants' stable and dynamic compositions.

	Stability	Balance	Liking	APB	DCM	VDCM
Dynamics	**−0.6750*****	**−0.5030*****	**0.1883****	**−0.2219*****	**−0.2816*****	**0.3733*****
Stability	-	**0.6292*****	**0.1014***	**0.2568*****	**0.2938*****	**−0.3464*****
Balance	-	-	0.0672	**0.2293*****	**0.2794*****	**−0.3373*****
Liking	-	-	-	**0.1048***	0.0374	−0.0384

APB, DCM, and VDCM were computed as described above. Significance is indicated as follows $p < .05$* $p < .001$** $p < .0001$***. Shown are the correlations between APB and DCM (VDCM) scores and the four perceptual judgments, across the two tasks including all compositions. In the overall correlation analysis of self-compositions, ratings and APB and DCM scores for stable and dynamic compositions for each observer were included, resulting in 20 (10 artworks for each 2 composition tasks) x 21 (observers) pairs of points (i.e. $N = 420$) per correlation. We find overall positive correlations between liking and stability, liking and dynamics, balance and stability, between DCM and balance, negative correlations between balance and dynamics, stability and motion (dynamics), and no significant correlation between liking and balance, and DCM scores and liking. Overall, the correlations of perceptual measures with the DCM and APB scores reached significance, suggesting that they indeed capture aspects of how balanced, dynamic and stable these compositions are perceived by our participants. DCM = deviation of the center of mass; VDCM = vertical deviation of the center of mass; APB = assessment of preference for balance.

Table A3. Mean ratings for original Suprematist artworks.

	M1	*M2*	*M3*	*P1*	*M4*	*M5*	*P2*	*K1*	*M6*	*M7*
										
$M_{dynamics}$	0.614	0.436	0.587	0.67	**0.325**	0.373	0.753	0.594	**0.788**	0.746
$SD_{dynamics}$	0.272	0.268	0.244	0.208	0.275	0.387	0.216	0.260	0.129	0.258
$M_{stability}$	0.326	0.439	0.376	0.328	0.459	**0.559**	0.302	0.305	0.221	**0.137**
$SD_{stability}$	0.249	0.249	0.259	0.210	0.305	0.369	0.273	0.231	0.211	0.136
$M_{balance}$	0.838	0.860	0.828	0.801	0.792	0.822	0.749	0.749	0.779	0.711
$SD_{balance}$	0.251	0.228	0.232	0.287	0.332	0.235	0.304	0.310	0.323	0.304
M_{liking}	0.405	0.484	0.478	0.515	0.395	0.491	0.626	0.441	0.506	0.474
SD_{liking}	0.265	0.228	0.261	0.300	0.302	0.276	0.344	0.250	0.283	0.324

Shown are mean ratings and corresponding standard deviations. Highlighted in bold are the most and least dynamic rated artworks, and the most and least stable rated artworks. For these conditions we assess participant consistency in preference. See main text.

Table A4. Correlations between preference for balance (APB) and DCM and ratings of the 10 original Suprematists' artworks.

	Stability	Balance	Liking	APB	DCM	VDCM
Dynamics	**−0.4763***** (−0.6750***)	**−0.3418***** (−0.5030***)	0.1303 (0.1883**)	**0.2528*** (−0.2219***)	0.0279 (−0.2816***)	−0.1036 (0.3733***)
Stability	-	**0.4917***** (0.6292***)	**0.2733**** (0.1014*)	**−0.1432*** (0.2568***)	−0.0993 (0.2938***)	0.0157 (−0.3464***)
Balance	-	-	0.0964 (0.0672)	**−0.1837*** (0.2293***)	−0.0879 (0.2794***)	0.0264 (−0.3373***)
Liking	-	-	-	0.0241 (0.1048*)	−0.1065 (0.074)	−.0610 (−0.0384)

APB, DCM, and VDCM were computed as described above. Significance is indicated as follows $p < .05$* $p < .001$** $p < .0001$***. Blue indicates values that are consistent with results obtained from the self-arranged compositions, red indicates values that are inconsistent (sign and/or significance) with corresponding values in Table A2 (here reprinted as a black value in parentheses). Compared to self-arranged compositions we find stronger positive correlations between liking and stability, liking and APB indices, a somewhat weaker positive correlation between balance and stability, a weaker negative correlation between balance and dynamics and stability and dynamics. There was no significant positive correlation between liking and dynamics, and correlations between APB/DCM/VDCM indices and all ratings changed sign and/or became non-significant. DCM = deviation of the center of mass; VDCM = vertical deviation of the center of mass; APB = assessment of preference for balance.

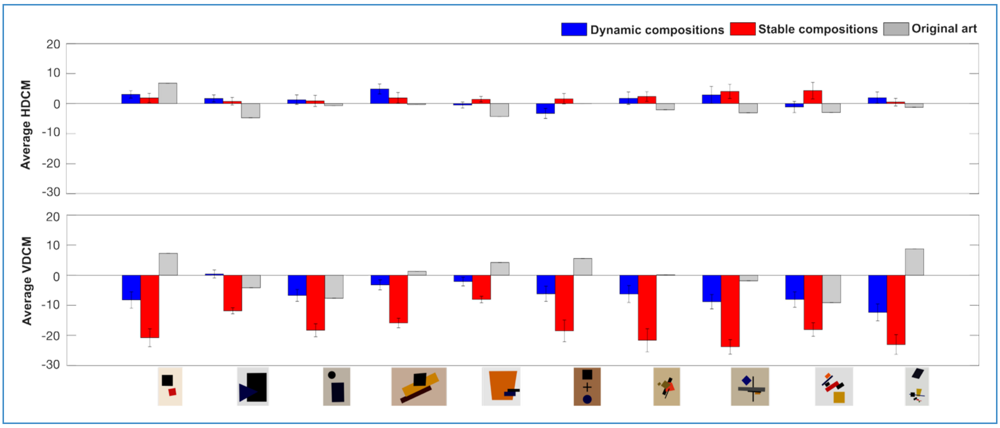


Figure A1. Horizontal and vertical deviations from the center of mass for self-arranged compositions and original artworks. The top panel shows the average deviation from the center of mass in the horizontal direction ((-) left to (+) right), the bottom panel average deviation from the center of mass in the vertical direction ((-) bottom to top (+)). For stable compositions (red bars) we see marked deviation towards the bottom half of the picture, since for this task participants tended to place elements into the lower half of the canvas. Also see Figure 4 in the main text. Deviations in the horizontal direction were not pronounced in either stable or dynamic compositions (blue bars). For original artworks deviations in the horizontal and vertical direction were not pronounced (gray bars).

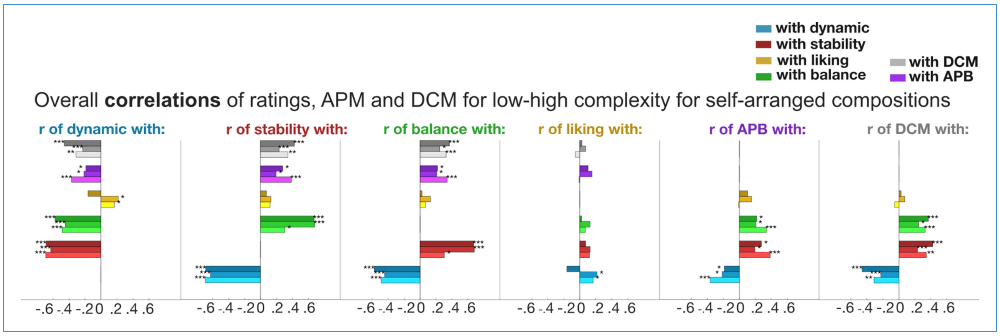


Figure A2. Correlations for self-arranged compositions across complexity levels. The graphical overview of correlations of ratings for self-arranged compositions illustrate how correlations change across complexity levels presented in Table 1 in the main text. The four rating tasks (how dynamic, stable, balanced, likable) and the two computational measures for balance (APB, DCM) are color coded: That is, correlations with dynamic ratings are shown in blue, correlations with stability ratings in red, balance ratings in green and liking in yellow. The three levels of complexity as determined by the number of elements in each composition are indicated by the position and by the brightness of the colors: The brightest color indicates a highest and the darkest the lowest complexity. Correlations are plotted as bars. Significance is indicated as follows $p < .05$* $p < .001$**$p <$ .0001***. Values of correlations correspond to values presented in Table 1 in the main text. DCM = deviation of the center of mass; APB = assessment of preference for balance.

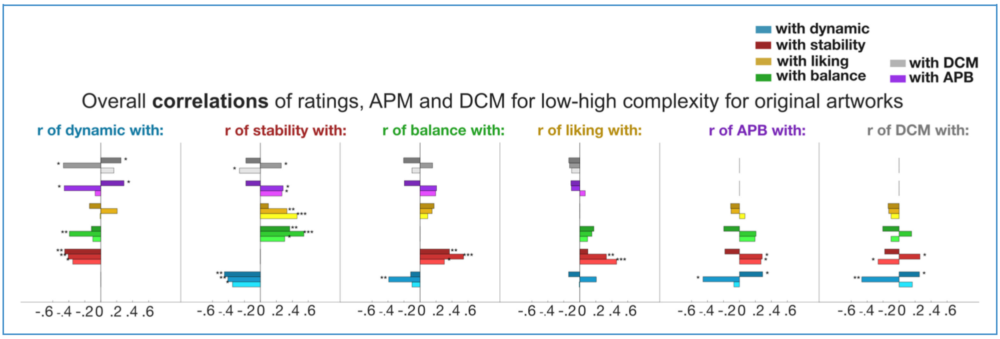


Figure A3. Correlations for original suprematist artworks. (A) The graphical overview of correlations of ratings for original artworks illustrate how correlations change across complexity levels. Correlations are plotted as bars, and correspond to the numerical values presented in Table 2 in the main text. The four rating tasks (how dynamic, stable, balanced, likable) and the two computational measures for balance: APB and DCM are color coded: That is, correlations with dynamic ratings are shown in blue, correlations with stability in red, balance ratings in green and liking in yellow. The three levels of artwork complexity are indicated by the position and by the brightness of the colors: The brightest colors indicate high—and the darkest colors the lowest complexity. Significance is indicated as follows $p < .05^{*}$ $p < .001^{**}$ $p < .0001^{***}$. Values of correlations correspond to values presented in Table 2 in the main text. DCM = deviation of the center of mass; APB = assessment of preference for balance.

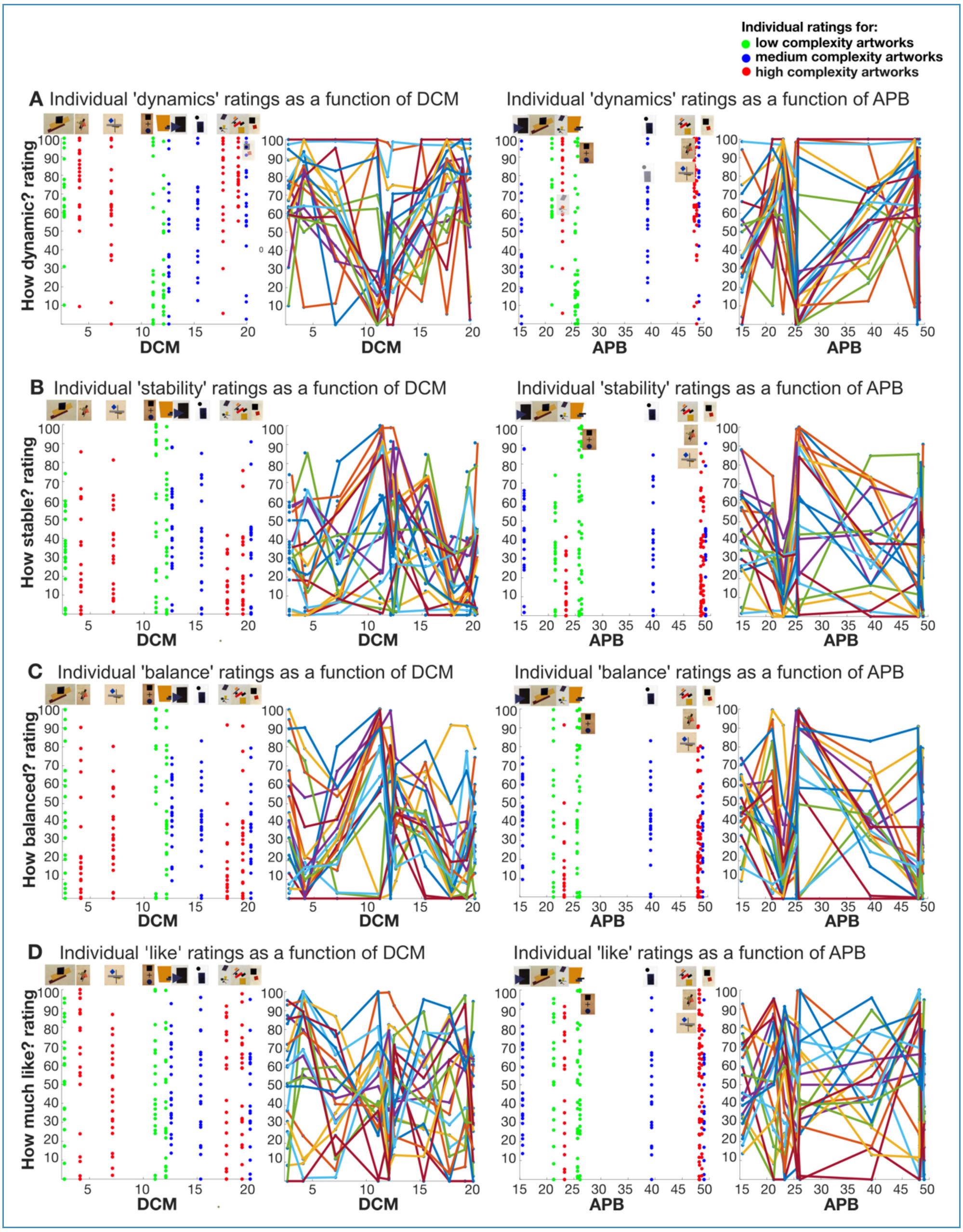


Figure A4. Rating results for original suprematist artworks. The correlations between APB/DCM and the ratings are mixed and often non-significant. To understand why that might be the case we created scatterplots that illustrate how overall ratings vary as a function of DCM and APB values and level of complexity. We also visualize individuals' rating as a function of artwork DCM/APB value (individuals are represented by a colored line that connects their ratings for the artworks). It is evident from these plots that there is much variability among participants in how a given artwork is judged, that the relationship between perceived attributes and DCM/APB changes as a function of complexity level, that within a complexity group (see Figure 2), DCM/APB values vary widely. (A**)** Shows results for "dynamic" ratings. **(B)** Results for "stable" ratings. **(**D) Results for "balance" ratings. **(**E) Results for "like" ratings. DCM = deviation of the center of mass; APB = assessment of preference for balance.

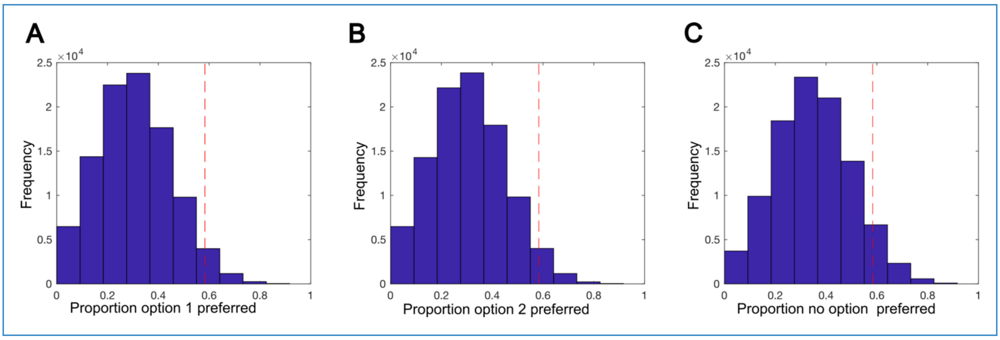


Figure A5. Simulation of an observer that makes "like" settings in a random fashion. For a given pair of options this observer sets its preferences according to a uniform random variable. If we take the difference between any 2 random preference settings, a certain proportion of the data will exceed the cutoff of 0.2 (if option 1 is preferred), −0.2 (if option 2 is preferred), and a certain proportion of the data will not exceed the cutoff (no preference). We simulated 12 such differences (as in the experiment, see text for details), noted the three proportions (1,2, no) and repeated this 100,000 times, yielding three distributions of proportions (A,B,C). Each distribution was then sorted and the value corresponding to the 95th percentile determined (vertical dashed red line). If in the actual data we find proportions larger than this cutoff value we reject the idea that the sample was generated by an observer that acted randomly. (A) Frequency distribution of the proportions of option 1 preferred. (B) Frequency distribution of the proportions of option 2 preferred. (C) Frequency distribution of the proportions of neither option preferred.

artistic feeling" and it was meant to be in sharp contrast to constructivism and materialism. In December 1915, the first actual exhibition of Suprematist paintings called "0.10" was shown in Petersburg including 35 paintings by Malevich. The first painting on Malevich's list in the catalog of the exhibition was an oil painting showing only a large black square on a quadratic white canvas (Figure 1, left). This painting represented the most radical abstract artwork painted so far and it became a landmark in the history of art. In his book "The Non-Objective World," published 1927 Malevich wrote: "In the year 1913, trying desperately to free art from the dead weight of the real world, I took refuge in the form of the square."