i-PERCEPTION a Pion publication

i-Perception (2013) volume 4, pages 57–77

dx.doi.org/10.1068/i0557 ISSN 2041-6695 perceptionweb.com/i-perception

Judging whether it is aesthetic: Does equilibrium compensate for the lack of symmetry?

Françoise Samuel
Faculté de Psychologie et des Sciences de l'Education, Université de Genève, Bd du Pont d'Arve 40, CH-1205 Genève, Switzerland; e-mail: Francoise.Samuel@unige.ch

Dirk Kerzel
Faculté de Psychologie et des Sciences de l'Education, Université de Genève, Bd du Pont d'Arve 40, CH-1205 Genève, Switzerland; e-mail: Dirk.Kerzel@unige.ch

Received 9 August 2012, in revised form 17 December 2012; published online 8 January 2013.

Abstract. In two experiments, we explored whether compositions made up of two or three rectangles received high aesthetics ratings when the composition was equilibrated, that is, when the center of mass of the weights represented by the areas of the rectangles were in the center of the composition. We supposed that equilibrated stimuli might be appreciated as much as symmetric stimuli. We further wanted to find out whether aesthetics ratings, balance ratings, and weight ratings differ from each other, and which stimulus characteristics influence each rating. We observed that the position of the center of mass of the compositions influenced the aesthetics ratings only slightly whereas it influenced the weight ratings more strongly. In contrast with this, the variation of the overall shape of the rectangles making up the compositions influenced the aesthetics ratings more strongly than the weight ratings. At first sight, balance ratings appeared intermediate between the aesthetics ratings and the weight ratings. However, when we varied the area ratio of the rectangles making up the composition, we observed that the balance ratings were independent of the two other ratings.

Keywords: aesthetics, composition, equilibrium, balance, symmetry, weight, ratings, visual perception.

1 Introduction

The present research aims to clarify the relation between aesthetic value, symmetry, equilibrium, and balance. Many studies have confirmed that symmetric compositions are judged to be more aesthetic, but it is unclear whether symmetry itself is judged more appealing or whether the high aesthetic value of symmetry arises from the perceived equilibrium of the composition. Indeed, if one interprets the darker areas of a composition as physical weights, then the center of mass of these weights is on the axis of symmetry, which corresponds to equilibrium. Asymmetric compositions may also reach physical equilibrium, for example, when small areas are moved further away from the center of the composition than large areas on the opposite side of the center. Thus, the question arises whether asymmetric compositions which are in equilibrium are judged to be as aesthetic as symmetric compositions. In other words, does physical equilibrium underlie effects of symmetry? Not separable from this query is the question whether the term "balanced" is equivalent to "equilibrated" like objects in the physical world, or whether it is equivalent to "harmonious." This secondary question is important to qualify previous studies that might or might not relate to equilibrium.

The preference for symmetric compositions was shown repeatedly (e.g., Campell, 1941; Lund & Anastasi, 1928). This strong preference in children decreases in adults, especially in adults with an education in arts (Frances, 1967; French, 1956). Analysis showed that renowned artworks have an overall symmetric structure, but with some carefully inserted asymmetry to avoid boredom and fixity (Gombrich, 1982; McManus, 2005). Asymmetry produces the impression of movement, as the observer tries in his mind to re-establish symmetry, as "man is born with an urge for stability and symmetry" (Gottlieb, 1958, p. 22). It must be mentioned that symmetry, especially mirror symmetry with respect to a vertical axis, is detected very quickly and with great reliability (e.g., Barlow & Reeves, 1979; Wagemans, 1997; Wenderoth, 1994).

Does the preference for symmetry generalize into a preference for equilibrated compositions? Researchers in aesthetics claimed that artwork which was termed “balanced” was considered more aesthetic. Specifically, Arnheim (1943), as a Gestalt theorist, emphasized the importance of balance, as it depends on the composition as a whole. He stated that a “good shape,” according to the Gestalt principles, corresponds to the shapes that we find in the physical world; therefore, “the artist’s striving for balance is revealed as just one aspect of a universal tendency in nature” (p. 74). This quotation shows that Arnheim uses the term “balance” in art as something that has a correspondence in the physical world. Similarly, Locher, Gray, and Nodine (1996) define pictorial balance in the following way: “For visual art stimuli such as paintings a balanced composition results when structural elements are grouped or organized in such a way that their perceptual forces compensate one another. Stated another way, balance is achieved when the elements of a painting are pitted against each other about a balance center so that the parts seem anchored and stable” (p. 1419). Here too, terms are used which make us think that “balance” and “physical equilibrium” are considered as equivalent. The importance of balance in artwork is demonstrated in several experimental studies. The subjective positions of the balance centers of original and modified pictures show that even small modifications in the picture lead to shifts in the position of the perceived balance center (Locher et al., 1996). Original paintings are perceived as more balanced in comparison to paintings in which an element of the composition has been moved, and such judgments can be done after very short exposure times (Locher & Nagy, 1996). Students of fine arts tend to build their compositions such that the center of mass is near the center of the composition (Locher, Cornelis, Wagemans, & Stappers, 2001).

To determine in an experimental way whether speaking of “balance” in pictures is only a metaphor for some aesthetic principle related to harmony or whether balance in pictures is closely related to physical equilibrium, and therefore calculable in the same way, McManus, Edmondson, and Rodger (1985) asked participants to find the balance point between two rectangles of unequal sizes. The results showed large interindividual differences. All participants interpreted the larger area as being heavier; however, some participants did not sufficiently take account of the size difference between the rectangles; others took overmuch account of the difference. The researchers concluded that, though observers find a balance point in a painting or an abstract stimulus, their judgment is correlated to physical equilibrium through subjective scales.

Other studies aimed at relating subjective balance to objective balance. Locher, Stappers, and Overbeeke (1998) divided the areas of square compositions by eight axes meeting in the center of the composition. The measure of balance was highest if the black areas were partitioned equally on both sides of each axis. It is important to note that in this calculation, the distance of the black areas from the center of the composition is not taken into account, which makes this measure fundamentally different from a measure of physical equilibrium, in which the distance of the mass (here the black area) from the center of the composition is of foremost importance. Very similar to Locher and colleagues’ balance measure is Wilson and Chatterjee’s (2005) assessment of preference for balance (APB). The APB scores the proportion of black area in relation to the four main axes. Additionally, it scores whether the black areas are distributed equally in the outer and inner zones of the composition. This inner–outer score reflects homogeneity rather than physical equilibrium. While physical equilibrium defined by the position of the center of mass[1] could have been calculated in these studies, the measures which were chosen may have added to the unclear meaning of the term “balanced.”

In a study on design aesthetics, Bauerly and Liu (2006) used the position of the center of mass as an objective criterion for balance (i.e., equilibrium). They examined symmetry and balance ratings given to compositions made up of rectangles. They found high correlations between subjective balance and objective equilibrium, and high correlations between subjective symmetry and objective symmetry. The authors also assessed which displays received the highest aesthetics ratings. Clearly, the compositions with a vertical reflexive symmetry axis than those with a horizontal reflexive symmetry axis were most liked. The authors mention that participants gave lower aesthetic ratings to the totally equilibrated but not perfectly symmetric stimuli. They do not mention whether these equilibrated stimuli were still more appreciated than the unequilibrated ones.

Recent studies show a renewed interest in the question of balance. Gershoni and Hochstein (2011) used Japanese calligraphic characters as stimuli. Even with very short exposure times, they collected reliable subjective balance ratings. However, they found no correlation between these ratings and

[1] Applied to the context of a black-and-white composition, the center of mass is the average position of all the units of black area (e.g., the pixels) within the composition (or the weighted average position of the black areas forming the composition).

the balance (APB) scores from Wilson and Chatterjee (2005). This demonstrates that APB, which measures something other than balance in the sense of equilibrium, also does not predict subjective balance. Gershoni and Hochstein's study shows that stimulus characteristics which are not connected to physical equilibrium enter the subjective aesthetics ratings: The rating differed when the characters were rotated, which revealed that vertical and horizontal elements are more appreciated than diagonal elements. Other criteria established by Gestalt theorists played a role in the ratings. Calligraphic characters with good grouping and closure were rated higher. Another study on balance was made by McManus, Stöver, and Kim (2011). They challenged Arnheim's (1974) theory that aesthetic balance is based on physical equilibrium. More precisely, according to Arnheim, the position of the center of mass has to fall on a main axis of the composition area in order to be balanced. In McManus and colleagues' experiment, participants communicated their judgment by cropping the composition such as to make them look best. The positions of the centers of mass of the compositions were calculated by assigning a weight to each pixel according to its gray shade. Even compositions that were based on a display used by Arnheim were not cropped such that the resulting centers of mass were on one of the main axes of the picture. Leyssen, Linsen, Sammartino, and Palmer (2012) investigated at which distance from a first object a second object is placed within a frame such that the composition would look pleasant. They found that balance had almost no impact on aesthetic preference for the placement of the second object, except that they found a slight preference to place it centrally.

So, what do we know about balance? Some studies state that aesthetic balance is equivalent to a physical equilibrium measure, but either they do not submit it to testing or they compare subjective balance to objective measures of balance which cannot be considered measures of equilibrium (the axes measures: Locher et al., 1998; Wilson & Chatterjee, 2005) as they violate the physical principle that distance to the center matters. Some studies use measures that are based on a two-dimensional equivalent to physical equilibrium (e.g., Bauerly & Liu, 2006; McManus et al., 1985, 2011), that is, these measures are based on the position of the center of mass which is the average position of the pixels weighted by their gray value. These studies cast doubts on the fact that compositions constructed such as to satisfy physical equilibrium are considered balanced or are considered aesthetic.

The two experiments we report here aim to clarify whether an asymmetric but equilibrated composition is rated nearly as pleasing as a symmetric composition (which by necessity is equilibrated). Also, we tested whether asymmetric compositions that are equilibrated are rated as significantly more pleasing than asymmetric compositions that are not equilibrated. As a crucial side issue, we want to sort out whether participants use the term "balanced" as equivalent to "harmonious" or as equivalent to "equilibrated," in the physical sense. We also need to know whether observers are able to make an equilibrium judgment if unequivocally asked to do so. This allows sorting out whether observers fail to find an equilibrated composition aesthetic because they fail to see whether something is in equilibrium or whether they do simply not find equilibrium aesthetic. The present study has implications for understanding already published research on visual aesthetics and for writing out recommendations for future research.

To address these issues, we created compositions made up of rectangles positioned in a rectangular frame. These compositions were varied according to the presence or absence of symmetry and according to the degree of equilibrium, measured by the distance from the center of mass to the center of the composition. We submitted the compositions to three types of ratings. First, we asked the participant to make an aesthetics rating (is it aesthetic? is it nice?); second, we asked for a balance rating (is it balanced? Are the rectangles well distributed?); third, we asked the participants to make a physical equilibrium rating (because of the "weight" of the black areas, would the composition fall to the left or to the right? And if there is a heavier side, how much heavier is it?). For the three ratings, we assessed the effect of symmetry, the effect of equilibrium, and the effect of the general shape of the stimuli.

2 Experiment 1

2.1 Methods

2.1.1 Participants

Twenty-four undergraduate students (19 females, 5 males; aged 19–42 years, M = 24.3, SD = 6.7) participated to fulfill a course requirement. All reported normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. Concerning their knowledge in physics, none of the participants had any training beyond the basic high school education; as to their knowledge or practice in arts, four participants expressed a strong interest in art or do artwork during their leisure time. All participants gave their informed consent in accordance with the policies of the University of Geneva.

2.1.2 Apparatus and general settings

The stimuli were generated by the COGENT 2000 toolbox of MATLAB and were presented on a CRT screen with a 1280 × 1024 pixel resolution at a refresh rate of 75 Hz. The visible screen dimensions (36.9 cm × 29.5 cm) had been adjusted to equalize vertical and horizontal pixel size. The experimental room was dimly lit. The participants' movements were not restrained; the viewing distance from the screen was approximately 55 cm. The experimental sessions were individual.

2.1.3 Stimuli

The stimuli were made up of three rectangles that had their centers on a horizontal line (see examples in Figure 1). One rectangle was separated from the next by a one-pixel line (0.03° of visual angle). So the rectangles were perceived as separate but attached to each other. The three rectangles were black and were presented on a white background, delimited by a thin black frame. The middle rectangle was at an equal distance from the left and right edges of the screen. There were 128 different stimuli; all were symmetric about the horizontal axis (the upper half is a mirror reflection of the lower half). Some were also symmetric about the vertical axis, because the right rectangle was equal to the left rectangle, and these are the stimuli we term symmetric. Other stimuli were not symmetric about the vertical axis, because the right rectangle differed from the left rectangle. These are the stimuli we term asymmetric.

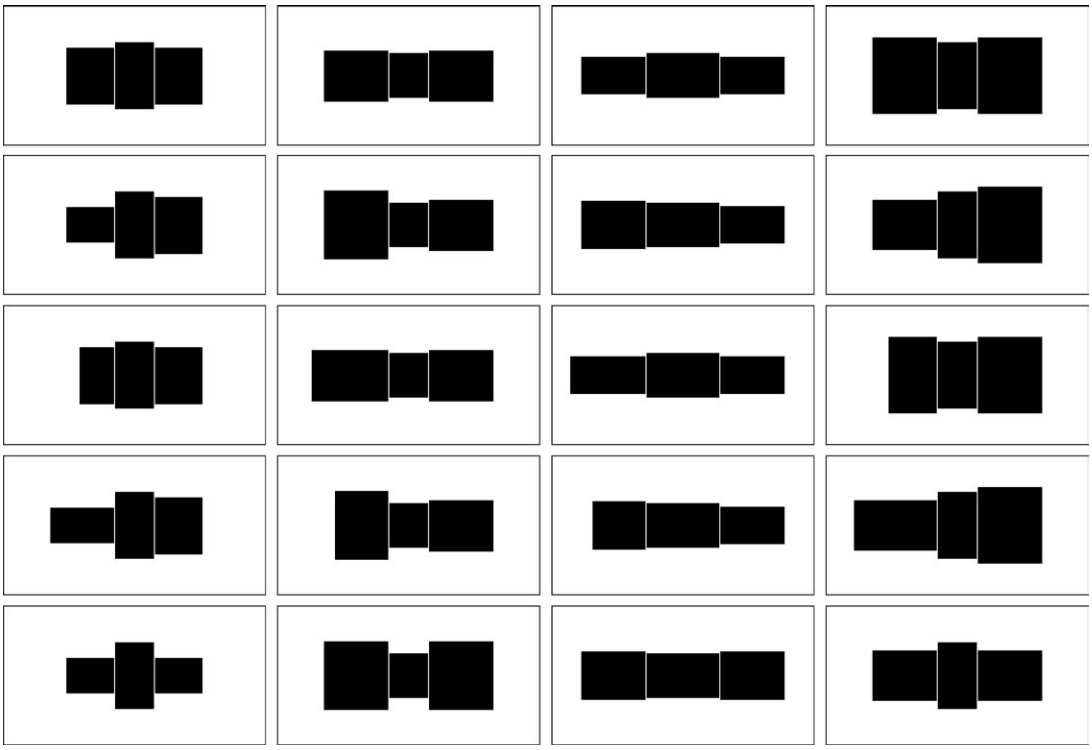

Figure 1. Example stimuli. The first line shows the initial symmetric stimuli. The second line shows the height-modified stimuli (positive or negative change of height in the left lateral rectangle) in comparison to the initial stimuli just above (e.g., first column, second line, the left rectangle is lower than just above). The third line stimuli are width-modified; they recover the initial height of the lateral rectangles, but change the width of their left lateral rectangle. For each stimulus, the width is calculated such that the position of the center of mass is the same than in the stimulus just above it. The fourth line shows equilibrated stimuli. Here the modified height was used, but the width was calculated such that the center of mass would be exactly in the center of the middle rectangle. In the fifth line, the stimuli are symmetric again, but now with the modified height (i.e., the same height as the left lateral rectangle on the second line). Concerning shape, the stimuli of the first column are all *mid-high* shaped; the stimuli of the second column are all *mid-low* shaped; the stimuli of the third and fourth columns change shape; you see the *stairs* shaped stimuli on the second and fourth lines. In the Aesthetics and Balance tasks, the three rectangles were presented with the surrounding frame, as visible in this figure.

The stimuli were constructed in the following way. An initial set of 16 symmetric stimuli was created. Two identical rectangles were positioned to the right and left of a middle rectangle such that the three centers were on a horizontal line. For half of these stimuli, the middle rectangle was higher than the lateral ones, and for the other half it was the contrary. Wide and narrow rectangles were equally represented. The overall width of the three rectangles taken together ranged from 480 to 720 pixels and the height, depending on the highest of the three rectangles, ranged from 136 to 276 pixels. See examples of symmetric stimuli in Figure 1 (first line).

By modifying the height of the left lateral rectangle of these initial stimuli, we built a first set of 16 asymmetric and unequilibrated stimuli. Half of the changes in height were positive in comparison to the initial height (ranging from +20% to +45%); half were negative in comparison to the initial height (ranging from −20% to −35%). The resulting heights of the left lateral rectangle ranged from 128 to 248 pixels. We avoided the modifications that would have given the lateral rectangle the same height as the middle rectangle. As the differences in height between the middle rectangle and the left lateral rectangle ranged from 16 to 112 pixels and the rectangles were vertically aligned, the steps from the middle rectangle to the lateral rectangle (i.e., half of the differences) ranged from 8 to 56 pixels (0.24° to 1.68° of visual angle). Examples of height-modified stimuli can be seen in Figure 1 (second line).

Then, we built a second set of 16 asymmetric and unequilibrated stimuli. For these stimuli, we changed the widths of the left lateral rectangles (their height was the same as for the initial symmetric stimulus). The change of width was chosen such that the displacement of the center of mass was equal to that caused by the modification of the height of the lateral rectangle as described in the above paragraph. That is, each stimulus with a change of width corresponded to another stimulus with a change of height, as their position of the center of mass was modified by the same amount. Figure 1 (third line) shows examples of width-modified stimuli.

We also built 16 asymmetric but equilibrated stimuli. They were similar to the asymmetric stimuli with the modified height of the left lateral rectangle, except that the width of the left rectangle was recalculated such that the three rectangles as a whole would be equilibrated.[2] That is, these stimuli would have the center of mass in the center of the middle rectangle and in the center of the screen. See examples of equilibrated stimuli in Figure 1 (fourth line).

To avoid an effect of stimulus side, we created mirror images of all the left side modified stimuli. For all the asymmetric stimuli, we made the corresponding right side modified stimuli. Thus, we had 32 more instances of asymmetric unequilibrated stimuli, and 16 more instances of asymmetric equilibrated stimuli.

Finally, we made a supplementary set of 16 symmetric stimuli. The sizes of their middle rectangles and the widths of their lateral rectangles were the same as in the initial set, but the lateral rectangle heights corresponded to those of the height-modified asymmetric stimuli (see examples in Figure 1, fifth line). It was important to have these stimuli to check that the modified heights elicited the same ratings as the initial heights. That is, if height-modified asymmetric or equilibrated stimuli were rated lower than the symmetric stimuli, we would be sure that it could not be due to a displeasing choice of height.

For the Aesthetics task and the Balance task, the background was delimited by a frame (for the Weight task, there was no frame). The frame was in a fixed position in the upper part of the screen and it was centered round the middle rectangle, which in itself was horizontally centered. The frame was 940 pixels wide and 500 pixels high (27.67° × 14.93° of the visual angle). For most stimuli, the three rectangles fitted into the frame without being tight. The maximum stimulus width was 803 pixels and its maximal height was 276 pixels (23.76° × 8.27° of the visual angle).

2.1.4 Characteristics of the stimuli

As already described, we created four types of stimuli, that is, stimuli were either symmetric (and consequently equilibrated), or they were equilibrated but not symmetric, or they were height-modified unequilibrated, or they were width-modified unequilibrated (the two unequilibrated types were of course not symmetric). There were 32 stimuli for each type of stimulus.

[2] When calculating the size and form of the altered lateral rectangle to obtain a specific position of the center of mass, for example an equilibrated stimulus, it is not sufficient to find an altered rectangle with the same area, but one also has to take into account the displacement of the center of mass of the lateral rectangle due to its change of form, as the position of the overall center of mass is the average position of the centers of mass of all rectangles weighted by their areas.

The strength of the disequilibrium also classifies the stimuli; we chose to classify the strength of the disequilibrium according to the eccentricity of the center of mass. The eccentricity is the distance in pixels between the center of the middle rectangle (and the frame, if there is one) and the center of mass of the three rectangles. By necessity, the symmetric and the equilibrated stimuli have the center of mass in the center of the composition, so the eccentricity of the center of mass is zero. The unequilibrated stimuli were classified as having a slight eccentricity (the centers of mass were 11–16 pixels eccentric) or a strong eccentricity (the centers of mass were 20–31 pixels eccentric). The slight and strong eccentricity classes contained 28 and 36 stimuli each. Due to the manner in which the stimuli had been constructed, the slight and strong eccentricity classes contained the same number of height-modified and width-modified stimuli.

Looking at the overall shape of the stimuli, we made another classification. Some stimuli have a middle rectangle which is higher than the lateral rectangles; others have a middle rectangle which is lower than the lateral rectangles. In the first case, we termed them *mid-high* stimuli; in the second case, we termed them *mid-low* stimuli. For some stimuli, the height of the three rectangles increased or decreased from left to right; in this case, we termed them *stairs* stimuli. The shape classification interacts with the symmetry-equilibrium classification. The *stairs* shape could occur only in equilibrated asymmetric stimuli and in height-modified stimuli. For these stimuli, there was a 50% occurrence of *stairs* stimuli with *mid-high* and *mid-low* stimuli occurring in 25% of cases each. For the symmetric and the width-modified stimuli, there were only *mid-high* and *mid-low* shapes, with a 50% occurrence of each. Figure 1 shows the different shapes: The stimuli of the first column are all *mid-high* and the stimuli of the second column are all *mid-low*. The stimuli of columns 3 and 4 have mixed shapes; for these columns, the second and fourth lines show *stairs* shapes.

2.1.5 The three tasks

The three tasks of the experiment were always run in the same order. First came the Aesthetics task, then the Balance task, last the Weight task. This order was mandatory because we wanted to see whether the aesthetics ratings were to some extent equilibrium ratings, and not the other way round. If we had asked for weight ratings first, we would have been unable to know whether an effect of physical equilibrium in the aesthetics ratings had been triggered by the weight ratings or whether aesthetics rating intrinsically take account of physical equilibrium. For the same reason, the Balance task had to be performed after the Aesthetics task. We had to be sure that the aesthetics ratings were uninfluenced by the balance ratings. Furthermore, it was important that the Balance task preceded the Weight task, in order to see, again, whether any physical equilibrium effect in the balance rating was intrinsic (i.e., not triggered by the weight ratings), or whether the balance rating was nearly equivalent to the aesthetics rating. For each task, the 128 stimuli were presented in random order.

In the Aesthetics task, participants had to rate how aesthetic or nice the composition was. We explained that the composition was the content of the frame, that is, the three rectangles in the area defined by the frame (Figure 2a). The rating was given by clicking on a broad vertical line with a length of about one-third of the screen width, which was located on the lower part of the screen, exactly in its horizontal center. The scale underlying the clickable line ranged from 0 (lower end of the line: not aesthetic) to 100 (upper end of the line: very aesthetic). All intermediate values could be recorded. There was no time constraint. The screen changed to the next stimulus as soon as the participant had clicked on the line.

In the Balance task, the screen presented the same elements as in the Aesthetics task: the composition and the vertical rating line. However, now the instructions were different. Participants had to rate how balanced, how well distributed the black areas were in the framed white area. The ratings could range from 0 (lower end of the line: not well distributed, not well balanced) to 100 (upper end of the line: very well distributed, very well balanced).

In the Weight task—which is how we termed the task concerned with physical equilibrium—the three rectangles were presented without the frame. They were in the same position they had been when surrounded by the frame, that is, they were on the upper part of the screen; as in the preceding tasks, the middle square was left–right centered. Under the center of the middle rectangle was a triangle (Figure 2b). The triangle's upper tip was in contact with the center of the lower edge of the middle rectangle, so that the three rectangles seemed to be supported by the triangle. Under the triangle was the response line, which in this task was horizontal. The instructions were that the participants should imagine that the black areas were weights and they should judge which side was heavier and how

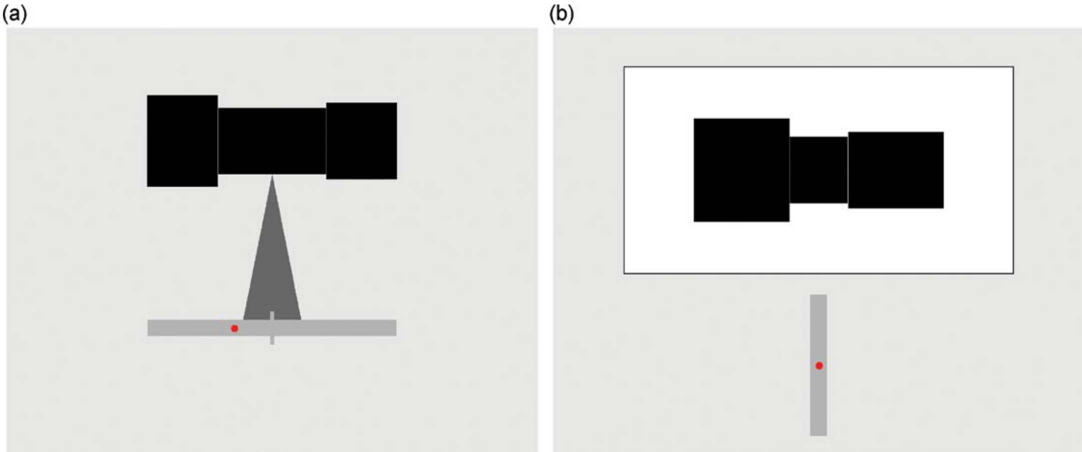

Figure 2. (a) Display for the Aesthetics and Balance tasks: the scale for the ratings is vertical. The position of the mouse was visible as a red dot. The top of the line was for high aesthetic or high balance ratings, the bottom for low aesthetic or low balance ratings. (b) Display for the Weight task: the central rectangle is supported by a triangle; the scale for the ratings is horizontal. Participants clicked on the scale to indicate on which side of the triangle, and to which extent the weight represented by the black areas would make the three rectangles fall to the left or right.

much heavier. The scale underlying the clickable line ranged from −100 (much heavier on the left side) to +100 (much heavier on the right side); a click on the center of the line corresponded to the judgment that both sides had the same weight. For data treatment, the scale of this third task was made compatible with the one of the two other tasks.

Before starting the experiment, the participants were told that they had to go through three tasks, but they were informed of the topic of the first task only. Participants were familiarized with the Aesthetics task in a small practice block. We did not give any feedback about the responses. If participants asked how to solve the task, we answered that it should be done in an intuitive way. As the Balance task differed only in instructions from the first task, participants were not asked to go through practice trials before starting the second task, but we made sure that the change of instructions was well understood. The Weight task differed more markedly from the two preceding tasks, so we asked participants to go through a short practice block before starting this new task.

After each task, we asked participants to fill in part of a questionnaire, to find out whether the participants had used a strategy or criterion to solve each task. As final questions, we asked participants whether they had special expertise in arts or physics.

2.1.6 Making the data of the three tasks compatible

The Aesthetics and the Balance tasks had 0–100 scales, and the Weight task had a −100 to +100 scale. To make the data of the Weight task comparable with the other two tasks, we computed the *absolute* values of the judgments and *inverted* the 0–100 polarity to have a scale which indicates 0 for "weight is eccentric" and 100 which indicates "weight is centered." Before doing this operation, we checked that taking the absolute values was not detrimental to the structure of the data, as taking the absolute values erases directional errors. Over all participants, we found less than 2% of directional errors (leftward answers, although the center of mass was to the right, and conversely). These data points were excluded from the analysis, as taking them into account would have incorrectly translated the responses of the participant.

2.1.7 Hypotheses

We expected that the three tasks would produce differing results. Consequently, we expected low between-task correlations. More specifically, we expected the between-task correlations to be quite lower than equivalent within-task correlations. Although we expected differences between the three tasks, we undertook the same analyses as if we had the same hypotheses for the three tasks (because not doing so would have biased the results too). Note that "higher ratings" mean "more aesthetic," "more balanced," or "weight is more centered" according to the task. In each task, we looked at the effect of stimulus type (symmetric, equilibrated, height-modified, width-modified), at the effect of eccentricity of center of mass (slight, strong) within the height- and width-modified stimuli, and at the effect of shape (*mid-high*, *mid-low*, *stairs*). We supposed that the effect of stimulus type, specifically, that the symmetric and equilibrated stimuli would receive higher ratings than the unequilibrated stimuli, would be strongest in the Weight task, as judging equilibrium was the explicit aim in this task. However, an effect of stimulus type was certainly expected in the Aesthetics task, as previous research had shown a preference for symmetric stimuli. We expected that stimuli with stronger disequilibrium would receive lower ratings than those with a slighter disequilibrium, specifically in the Weight task, again as this was explicitly the task ("show me how much more heavy it is on the heavier side"). For self-evident reasons, we hypothesized that the effect of shape would be strongest in the Aesthetics task; we expected the *mid-high* shapes to receive the highest ratings, as this shape is most rounded, and this seems optimal as a Gestalt, and we expected the *stairs* shape to get the lowest ratings as it is most obviously asymmetric with respect to a vertical axis. We thought that shape would not influence the weight ratings as shape is varied independently of equilibrium. We expected the Balance task to produce results that are intermediate between the two other tasks.

2.2 Results

2.2.1 Do the results of the three tasks differ?

We calculated the within- and between-task correlations, as low between-tasks correlations can be interpreted as differences between the tasks only in contrast with high within-task correlations. For the within-task correlations, each asymmetric stimulus was paired up with its mirror-symmetric counterpart and each symmetric stimulus was paired up with a similar symmetric stimulus (i.e., each symmetric stimulus of the initial series was paired up with the corresponding stimulus of the series with the modified height). Accordingly, the within-task correlations were based on 64 pairs. Likewise, for the between-task correlations we took the same pairs, but across tasks. That is, from one task we took one half of the pair and from the other task the other half of the pair. Each task furnished in one comparison the first pair, and in the other comparison the other pair of stimuli. Consequently, the between-task correlations were based on the same 64 pairs of stimuli than the within-task correlations.

The mean Aesthetics within-task correlation in ratings was 0.45 (individual correlations ranged from −0.05 to 0.84; five individual correlations were below 0.30). Within the Balance task, the mean correlation was 0.70 (individual correlations ranged from 0.37 to 0.91). The mean Weight within-task correlation was 0.63 (individual correlations ranged from 0.34 to 0.89). That is, of the three tasks, the Aesthetics task produced the least reliable results and the Balance task produced the most reliable results (still, about half of the variance is unexplained).

The mean correlation between the Aesthetics task and the Balance task was 0.38 (individual correlations ranged from −0.49 to 0.81; seven individual correlations were below 0.30). The mean correlation between the Balance and Weight tasks was 0.48 (individual correlations ranged from 0.17 to 0.76; four were below 0.30). Last, the mean correlation between the Aesthetics and Weight tasks was 0.28 (individual correlations ranged from −0.34 to 0.61; eleven were below 0.30). In each case, several participants have correlations near zero or in the opposite than expected direction. We found two participants with coefficients of 0.50 or above concurrently in the three between-task correlations.

After carrying out a variance-stabilizing Fisher transformation of the correlation coefficients, we tested whether each mean within-task correlation was larger than the two mean between-task correlations involving this task. In one case, the t-test was not significant: The mean within-task correlation in the Aesthetics task was not higher than the mean correlation between Aesthetics and Balance tasks, $t(23) = 1.32, p = 0.20$. In the five other cases, the t-tests showed that the within-task mean correlations were significantly higher than the mean between-task correlations (for the 5 remaining comparisons: $t(23) \geq 3.83, p \leq 0.001$). That is, generally, the results differ from one task to another.

2.2.2 Effect of stimulus type

In the Aesthetics task, the repeated-measure ANOVA showed an effect of stimulus type on the ratings (75.8, 50.9, 51.0, and 50.4, for symmetric, equilibrated, height-modified, and width-modified, respectively), $F(3, 69) = 25.42, p < 0.001$, and $\eta^2_p = 0.53$. The Helmert contrasts revealed that the symmetric stimuli received higher ratings than the three other types of stimuli, $F(1, 23) = 50.95$, $p < 0.001$, and $\eta^2_p = 0.69$. We found no difference between the equilibrated and the two types of unequilibrated stimuli, $F(1, 23) < 1$, $p = 0.937$, $\eta^2_p < 0.01$, and, within the unequilibrated stimuli, we detected no difference between the height- and width-modified stimuli, $F(1, 23) < 1$, $p = 0.877$, and $\eta^2_p < 0.01$. Please also refer to Figure 3.

In the Balance task, we also found an effect of stimulus type (86.6, 37.7, 37.8, and 52.6, for symmetric, equilibrated, height-modified, and width-modified, respectively), $F(3, 69) = 51.97$, $p < 0.001$, and $\eta^2_p = 0.69$. The Helmert contrasts showed that the symmetric stimuli received higher ratings than the three other types of stimuli, $F(1, 23) = 110.50$, $p < 0.001$, and $\eta^2_p = 0.83$. Within the asymmetric

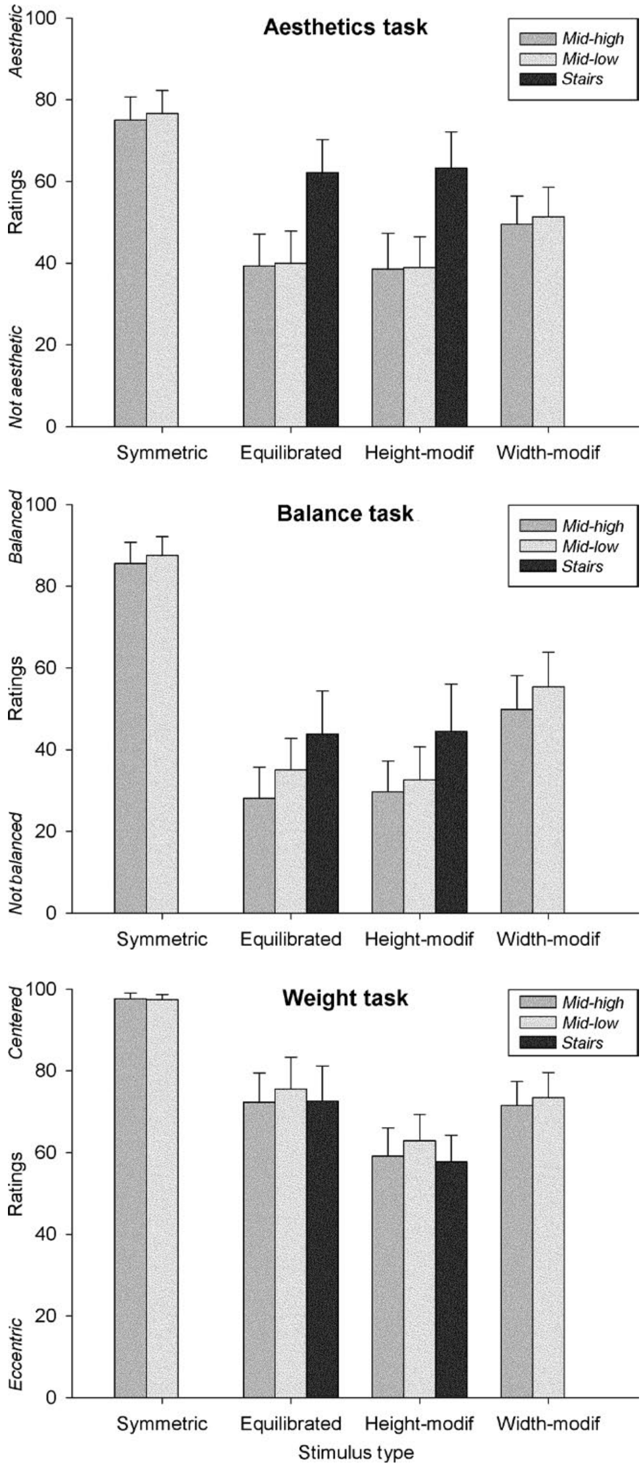

Figure 3. Results of the three tasks. For each task, the ratings are plotted according to stimulus type (symmetric, equilibrated, height-modified unequilibrated, width-modified unequilibrated) and stimulus shapes (*mid-high*, *mid-low*, *stairs*).

stimuli, we found a difference between the equilibrated and the two types of unequilibrated stimuli, however not in the expected direction: Equilibrated stimuli were rated lower than the unequilibrated stimuli, $F(1, 23) = 7.90$, $p = 0.010$, and $\eta^2_p = 0.26$. Within the unequilibrated stimuli, the height-modified stimuli were rated lower than the width-modified stimuli, $F(1, 23) = 8.60$, $p = 0.008$, and $\eta^2_p = 0.27$.

The overall higher ratings in the Weight task in comparison with the two other tasks can be explained by the fact that raters tend to choose the middle section of a scale when they have no opinion; in the initial scale, this corresponded to ratings near 0. With the conversion to inverted absolute values, this corresponds to ratings near 100. These "no opinion" ratings raise all the means. However, differences between stimuli would not be influenced by this overall raise. The Weight task generated an effect of stimulus type (97.5, 73.3, 59.4, and 72.5, for symmetric, equilibrated, height-modified, and width-modified, respectively), $F(3, 69) = 60.95$, $p < 0.001$, and $\eta^2_p = 0.73$. The Helmert contrasts confirm that the symmetric stimuli get higher ratings than the three types of asymmetric stimuli, $F(1, 23) = 121.67$, $p < 0.001$, and $\eta^2_p = 0.84$. As expected, the equilibrated stimuli received higher ratings than the two types of unequilibrated stimuli, $F(1, 23) = 7.42$, $p = 0.012$, and $\eta^2_p = 0.24$. However, a look at Figure 3 shows that this difference is due to the low rating of the height-modified stimuli only. Within the unequilibrated stimuli, the width-modified stimuli received higher ratings than the height-modified stimuli, $F(1, 23) = 38.04$, $p < 0.001$, and $\eta^2_p = 0.61$.

2.2.3 Effect of disequilibrium

Concerning exclusively the unequilibrated stimuli, we tested the effect of slight versus strong disequilibrium. In the three tasks, we found, as expected, that those with a slighter eccentricity of center of mass were given higher ratings than those with a stronger eccentricity, Aesthetics task: 53.4 vs. 48.7, $t(23) = 3.67$, $p = 0.001$, $\eta^2_p = 0.37$; Balance task: 50.0 vs. 41.5, $t(23) = 4.53$, $p < 0.001$, $\eta^2_p = 0.47$; Weight task: 70.8 vs. 62.1, $t(23) = 9.23$, $p < 0.001$, $\eta^2_p = 0.79$.

2.2.4 Effect of shape

First, we ran ANOVAs that included only the *mid-high* and *mid-low* shaped stimuli, which are found in the four stimulus types. We wanted to know whether these two shapes receive different ratings and whether the shape interacts with the stimulus type (2 Shapes [*mid-high*, *mid-low*] * 4 Types [symmetric, equilibrated, height-modified, width-modified]). We will not mention the effects due to stimulus type, as we reported them above. The repeated-measure ANOVA showed no difference between *mid-high* and *mid-low* in the Aesthetics task (54.5 vs. 55.8), $F(1, 23) < 1$, $p = 0.415$, $\eta^2_p = 0.03$, and no interaction between shape and stimulus type, $F(3, 69) < 1$, $p = 0.911$, $\eta^2_p < 0.01$. In the Balance task, we found an effect of shape with lower ratings for the *mid-high* stimuli in comparison to *mid-low* stimuli (54.8 vs. 58.9), $F(1, 23) = 18.06$, $p < 0.001$, $\eta^2_p = 0.44$, but no interaction between shape and type, $F(3, 69) = 2.45$, $p = 0.070$, $\eta^2_p = 0.10$. In the Weight task, we also found lower ratings for the *mid-high* stimuli in comparison to the *mid-low* stimuli (78.4 vs. 80.2), $F(1, 23) = 6.82$, $p = 0.016$, $\eta^2_p = 0.23$. As in the two other tasks, there was no interaction between shape and stimulus type, $F(3, 69) = 1.33$, $p = 0.273$, $\eta^2_p = 0.06$. The effect of shape can also be observed in Figure 3.

Then, we ran ANOVAs including only the equilibrated and the height-modified stimuli, as they are the stimulus types containing *stairs* shaped stimuli (2 Types [equilibrated, height-modified]) * 3 Shapes [*stairs*, *mid-high*, *mid-low*]). In the Aesthetics task, the ANOVA showed a highly significant effect of shape, $F(2, 46) = 21.96$, $p < 0.001$, $\eta^2_p = 0.49$, and no interaction between shape and stimulus type, $F(2, 46) < 1$, $p = 0.618$, $\eta^2_p < 0.10$. The Helmert contrast showed that the *stairs* shape received significantly higher ratings in comparison to the *mid-high* and *mid-low* stimuli taken together (62.7 vs. 39.0 and 39.5), $F(1, 23) = 23.81$, $p < 0.001$, $\eta^2_p = 0.51$. In the Balance task, the ANOVA showed a significant effect of shape, $F(2, 46) = 5.28$, $p = 0.009$, $\eta^2_p = 0.19$, and no interaction between shape and stimulus type, $F(2, 46) = 2.30$, $p = 0.112$, $\eta^2_p < 0.10$. The contrast showed that the *stairs* shape were rated significantly higher than the two other shapes (44.1 vs. 28.8 and 33.9), $F(1, 23) = 5.03$, $p = 0.035$, $\eta^2_p = 0.18$. In the Weight task, the ANOVA showed a significant effect of shape, $F(2, 46) = 3.82$, $p = 0.029$, $\eta^2_p = 0.14$, and no interaction between shape and stimulus type, $F(2, 46) < 1$, $p = 0.516$, $\eta^2_p < 0.10$. The contrast showed that the effect of shape was not attributable to a difference between the *stairs* shapes and the two other shapes (65.3 vs. 65.8 and 69.3), $F(1, 23) = 2.84$, $p = 0.105$, $\eta^2_p = 0.11$.

2.2.5 Checking whether the modification of height in itself influences the ratings

We checked whether we could observe a difference in ratings between the initial symmetric stimuli and the symmetric stimuli after the modification of height (compare the first and fifth lines of Figure 1). This was not the case (Aesthetics task: 75.1 vs. 76.6; Balance task: 86.4 vs. 86.7; Weight task: 97.6 vs. 97.4, for initial and modified, respectively). Consequently, the height modification in itself does not lower the ratings.

2.3 Discussion of Experiment 1

Based on the review of the literature (e.g., Bauerly & Liu, 2006; Lund & Anastasi, 1928; McManus, 2005), we foresaw that symmetric stimuli would get highest ratings in the three tasks: Symmetric stimuli were considered most aesthetic, most balanced, and having the weight most centered. We examined individual results for exceptions to the aesthetic preference for symmetry. Only one participant rated the symmetric stimuli aesthetically lower than the asymmetric ones; this was not due to the fact that this participant was not able to perceive symmetry, as he gave highest ratings to symmetric stimuli in the Weight and Balance tasks.

Our hypothesis that the equilibrated stimuli would receive ratings that were nearly as high as the symmetric stimuli and certainly higher than the unequilibrated stimuli was falsified in the three tasks. Even in the Weight task, where we found some difference in favor of the equilibrated stimuli, this advantage held only over the height-modified stimuli, and not over the width-modified stimuli. However, in the three tasks, there is an effect of the strength of disequilibrium within the unequilibrated stimuli: stimuli with a stronger eccentricity of the center of mass receive lower ratings. As anticipated, this effect was more marked in the Weight task, confirming that observers perceived the equilibrium rather well. We suppose that the inconsistent effect of eccentricity in the three tasks (no difference in ratings between zero eccentricity and slight eccentricity, but a decrease in ratings between slight and strong eccentricity) could be due to the fact that with our stimuli, the equilibrated stimuli are less symmetric than the unequilibrated stimuli. The equilibrated stimuli are asymmetric because of the height *and* the width of one lateral rectangle in comparison to the other, whereas the unequilibrated stimuli are asymmetric because of the height *or* the width. Therefore, we suppose that equilibrated stimuli fail to receive higher aesthetics, balance, and weight ratings because they are less symmetric than unequilibrated ones. On the other hand, within the unequilibrated stimuli, asymmetry increases with increased eccentricity. Thus, our results might reflect that the effects of equilibrium and symmetry are confounded.

For a few participants the ratings in the different tasks were correlated. However, for most participants the correlations from one task to the other were low, which prepared us to find different patterns of responses according to tasks. The analysis of the effect of shape gave unforeseen results. The difference between *mid-high* and *mid-low* shapes did not appear in the aesthetics ratings where such an effect would have seemed least surprising, but it appeared unexpectedly in the balance and weight ratings: *mid-low* shapes seemed more balanced; their weight seemed more centered. Maybe the analogy between the *mid-low* shape and the shape of a barbell induced this effect. Alternatively, observers might have perceived that with the *mid-low* shape, the weight is more spread out, therefore better distributed (the term we used to paraphrase *balanced*). Furthermore, the high aesthetics ratings for the *stairs* stimuli came as a surprise. These high ratings can be explained by the fact that this shape, although obviously asymmetric about the vertical axis, emphasizes the symmetry about the horizontal axis. It might also show a preference for orderliness, that is, a repetition of increase or decrease. In the Balance task, the *stairs* shape keeps a slight advantage over the other shapes, and we have no explanation for this effect, except the analogy with the Aesthetics task. In the Weight task, there is no effect of the *stairs* shape.

If we look only at the *mid-high* and *mid-low* stimuli in the Aesthetics and Balance tasks (Figure 3), we find that the width-modified stimuli seem less penalized by their asymmetry than the two other asymmetric types of stimuli. The difference in height between the two lateral rectangles (when it is not in a progression as with *stairs* stimuli) seems to disrupt aesthetics and balance more than a change in width. This can be explained by the fact that a difference in height is perceived much more easily than a difference in width, because the first can be done by making one horizontal saccade from one lateral rectangle to the other, and the latter necessitates to compare two lengths in the peripheral regions of the stimulus. Therefore, it seems that asymmetric equilibrated objects are perceived as less aesthetic and less balanced than asymmetric unequilibrated objects with an asymmetry that is not easily perceived.

We undertook a follow-up experiment, essentially to disentangle the effect of equilibrium from the effect of symmetry. To this end, we designed series of stimuli for which the variation of symmetry and equilibrium was independent. Experiment 2 avoided the special situation where a middle rectangle indicates the center of the composition in redundancy with the frame. Rather, equilibrium was only defined relative to the center of the frame. To stay with most simple stimuli, we reduced the composition to two rectangles in a frame. On the one hand, we created equilibrated and unequilibrated stimuli that were equally bare of symmetry. On the other hand, we created symmetric rectangle designs that were centered or not centered. Therefore, symmetric designs positioned centrally made the composition as a whole equilibrated, while symmetric designs positioned off-center made the composition unequilibrated. We hypothesized that for equally symmetric stimuli, observers would rate equilibrated stimuli higher than unequilibrated ones.

Furthermore, we wanted to explore the effect of area ratio of the two rectangles making up the composition, a characteristic that is only indirectly related to equilibrium. We hypothesized that ratio, as it had been the case with shape in the first experiment, would influence the ratings in different ways according to task.

3 Experiment 2

3.1 Methods

3.1.1 Participants

Twenty-four undergraduate students (22 females, 2 males; aged 19–29 years, $M = 21.3$, $SD = 2.7$) participated in completion of a course requirement. All reported normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. Concerning their knowledge in physics, two participants majored in physics from high school. Concerning their knowledge or practice in arts, four participants expressed an interest in art or majored in arts from high school. All gave their informed consent in accordance with the policies of the University of Geneva.

3.1.2 Apparatus and general settings

The technical specificities were the same as in Experiment 1.

3.1.3 Stimuli

The stimuli consisted of two rectangles that had their centers on a horizontal line (see examples in Figure 4), that is, all stimuli were symmetric about the horizontal axis. The two rectangles were black and they were presented on a white background, delimited by a thin black frame, which was horizontally centered on the screen.

The stimuli were constructed in the following way. One of the rectangles always had an area of 60,000 square pixels and the other rectangle had an area of 60,000, or 45,000, or 30,000, or 15,000 square pixels, corresponding to areas ranging from 50 to 12.5 cm^2. For a rectangle of a chosen area, we determined the width as the square root of the area times a random step factor (varying in the range from 0.71 to 1.40) and the height as the area divided by the width. For example, for a rectangle with an area of 30,000 square pixels, and a random step factor of 1.27, the rectangle has a width of 220 pixels ($\sqrt{30{,}000} \times 1.27$) and a height of 136 pixels (30,000/220). With this mode of calculation, the extent of deformation of the squares into high and wide rectangles was the same. For the smallest area, the shortest possible rectangle side length was 87 pixels (2.6° of the visual angle) and for the biggest area the longest possible side length was 343 pixels (10.3° of the visual angle). The random procedure was done independently for the right and left rectangles. The randomization was independent for each participant. When the ratio of the areas was not 1:1, the smaller area was equally often on the left or right, which resulted in the seven ratio situations 1:1, 1:¾, 1:½, 1:¼, ¾:1, ½:1, and ¼:1. The two rectangles were separated from each other by a gap which we varied systematically. Its width was of 10, 20, 30, or 50 pixels (0.3°–1.5° of the visual angle).

We varied the position of the center of mass of the two rectangles (taken together; taking account of the gap) such that the positions of the centers of mass were 0, −60, +60, −120, and +120 pixels eccentric toward the left or right (0°, ±1.8°, and ±3.6°) in comparison to the center of the frame. Note that the distances from the center of mass to the center of composition are much bigger than in Experi-

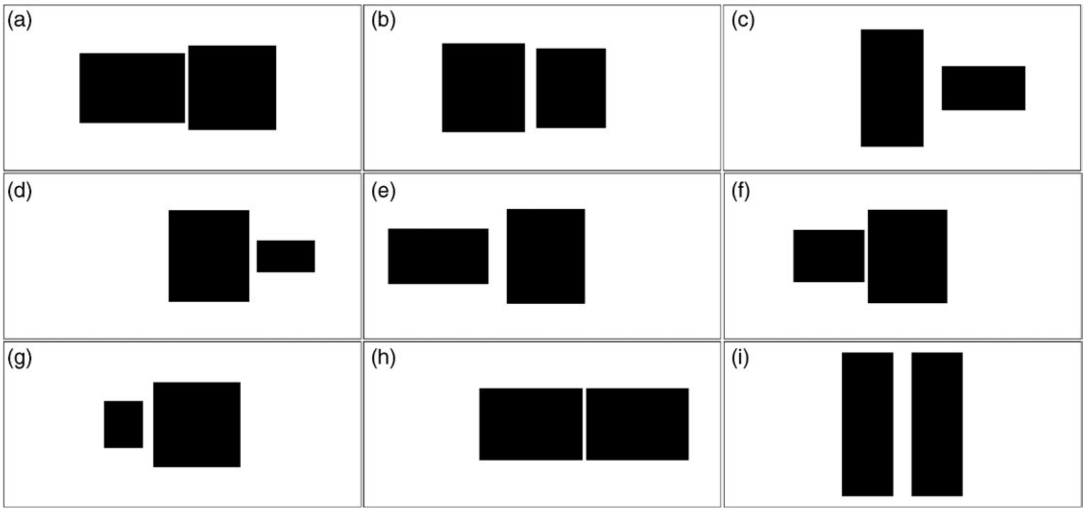

Figure 4. Examples of stimuli. In the three parts of the experiment, the two rectangles were presented with the surrounding frame, as visible in this figure (only one at a time). Description of the nine stimuli (left to right and from top to bottom) in terms of area ratio, gap between the rectangles, and eccentricity of the center of mass in comparison to the center of the frame: (a) 1:1, 10, 0; (b) 1:¾, 30, −60; (c) 1:½, 50, +60; (d) 1:¼, 20, +120; (e) ¾:1, 50, −120; (f) ½:1, 10, −60; (g) ¼:1, 30, 0; (h) 1:1, 10, +120; (i) 1:1, 50, 0. The stimuli (h) and (i) are from the supplementary set, that is, they are identical-rectangles stimuli.

ment 1, but here the central reference is much weaker, as there is no central rectangle, so the deviations from the center might be about as perceptible as in Experiment 1. Counterbalancing the seven area ratios, the four gaps, and the five eccentricities corresponded to 140 stimuli.

In addition to these 140 dissimilar-rectangles stimuli, we created an extra 12 stimuli with two identical rectangles (as this could only have happened by chance on rare occasions). For these identical-rectangles stimuli, the eccentricity of the position of the center of mass was zero in four cases (so here the display was entirely symmetric), plus or minus 60 pixels in four cases, plus or minus 120 pixels in four cases. The shapes of the rectangles (both rectangles of the stimulus equally wide and equally high) and distances between the rectangles were maximally varied. Altogether (dissimilar- and identical-rectangles stimuli), there were 152 stimuli.

The two rectangles were positioned in a white area, 1016 pixels wide and 468 pixels high (29.8° × 13.9° of the visual angle), framed by a thin black line. This framed area was the reference for the eccentricity of the center of mass (a stimulus with zero eccentricity would have its center of mass at the center of the frame). For most stimuli, the two rectangles fitted into the frame without being tight.

3.1.4 The three tasks

The three tasks of the experiment were the same as in Experiment 1 and they were run in the same order. However, there was a difference in the display of the Weight task. In Experiment 2, there was a frame around the stimuli in the three tasks (and not only in the Aesthetics and Balance tasks; Figure 5b; compare with Figure 2b). Contrary to Experiment 1, where the stimuli had an intrinsic center, in Experiment 2 the frame, and nothing else, determines the center of the composition. Therefore, the frame must be present for the task to make sense. Also, we wanted to make sure that the rating concerned the same display in each task. Accordingly, in Experiment 2, the triangle's upper tip was in contact with the center of the lower edge of the frame, so that the frame seemed to be supported by the triangle. Under the triangle was the horizontal response line. The instructions were that the participants should imagine that the black areas in the frame were weights and they should judge which side of the frame was heavier and how much heavier. In all else, the tasks were performed as in Experiment 1.

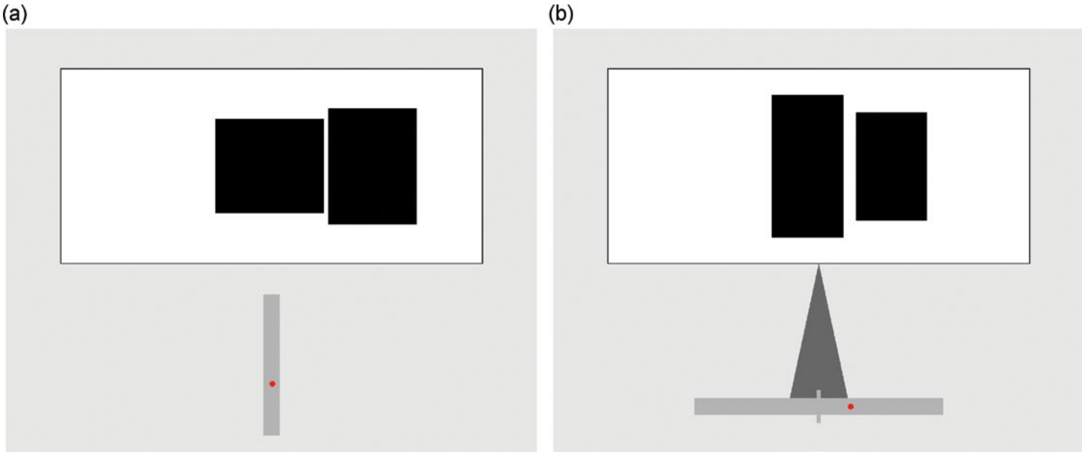

Figure 5. (a) Displays for the Aesthetics and Balance tasks. In the first task, participants clicked on the vertical line to indicate their aesthetics ratings; in the second task, they clicked to indicate their balance ratings. (b) Display for the Weight task. The frame is propped on the triangle. Participants clicked on the horizontal line to indicate on which side, and to which extent, the weight represented by the black areas would make the frame fall to the left or right.

3.1.5 Making the data of the Weight task comparable to the data of the two other tasks

We treated the data of the Weight task as those of the Weight task in Experiment 1. Before computing the absolute values and inverting the 0–100 polarity, we found, as in the preceding experiment, less than 2% of responses with directional errors. We excluded these data points from the analysis.

3.1.6 Hypotheses

On the basis of the results of the previous experiment, in the three tasks we expected that identical-rectangles stimuli would receive higher ratings than the dissimilar-rectangles stimuli, as the two identical rectangles are symmetric in themselves (and if they are centered, then the whole composition is symmetric). Furthermore, we expected an effect of position of center of mass, that is, we expected that the ratings would decrease from equilibrium to disequilibrium, and more so for strong than for slight disequilibrium, especially in the Weight task, as equilibrium is the explicit criterion in this task. A point of main interest is the comparison between the identical- and dissimilar-rectangles stimuli in the situation where the center of mass is in the center of the composition, as this is a comparison between symmetry and equilibrium.

Regarding the variation of ratio, we supposed that it should have the strongest effect on the aesthetics ratings, because it is concerned with the shape of the stimuli. Knowing that symmetry is most appreciated, we supposed that stimuli with the 1:1 ratio would be more appreciated than stimuli combining a large-area rectangle with a small-area rectangle, as the latter are more obviously asymmetric. We supposed that the variation of ratio would have no effect on the weight ratings as ratio is varied independently from equilibrium.

3.2 Results

3.2.1 Do the results of the three tasks differ?

The mean Aesthetics within-task correlation was 0.25 (individual correlations ranged from −0.16 to 0.56; 14 individual correlations were below 0.30). Within the Balance task, the mean correlation was 0.46 (individual correlations ranged from −0.02 to 0.85; six were below 0.30). The mean Weight within-task correlation was 0.54 (individual correlations ranged from 0.25 to 0.73; one was below 0.30). That is, the ratings varied very much between two similar stimuli. With an absence of correlation for many participants, the Aesthetics task produced the least reliable results. Note that in this experiment the shape of stimuli varied randomly (within constraints); therefore, pairs are matched according to mirror characteristics but not according to exact mirror shape; this partially explains the low correlations.

The between-task correlations were based on the corresponding 76 pairs of stimuli. The mean correlation between the Aesthetics and Balance tasks was 0.12 (individual correlations ranged from −0.24 to 0.55; 20 individual correlations were below 0.30). The mean correlation between the Balance and Weight tasks was 0.28 (individual correlations ranged from −0.14 to 0.70; 14 were below 0.30). Last, the mean correlation between the Aesthetics and Weight tasks was 0.09 (individual correlations ranged from −0.22 to 0.50; 22 were below 0.30). None of the participants had coefficients of 0.50 or above concurrently in the three between-task correlations.

After carrying out a Fisher transformation of the correlation coefficients, we tested whether each mean within-task correlation was larger than the two mean between-task correlations involving this task. All six comparisons showed significantly higher within- than between-task correlations, $t(23) \geq 3.01$, $p \leq 0.006$.

3.2.2 The effect of identicalness and the effect of eccentricity of center of mass

Figure 6 (left) shows the results concerning the eccentricity of the center of mass for the three tasks, separately for the stimuli made of two identical rectangles (dark gray) and for those made of two dissimilar rectangles (light gray). Within each task, we ran ANOVAs testing the difference between identical- and dissimilar-rectangles stimuli and the effect of center of mass eccentricity (2 Identicalness [identical, dissimilar] * 3 Eccentricities [0, ±60, ±120]). Before pooling the data of the −60 eccentricity with the data of the +60 eccentricity, and −120 with +120, we checked for differences between these conditions, but in none of the tasks did we find any difference attributable to leftward versus rightward eccentricity. In the Aesthetics task, we found that identical-rectangles stimuli were rated higher than dissimilar-rectangles stimuli (64.4 vs. 49.6), $F(1, 23) = 6.05$, $p = 0.022$, $\eta^2_p = 0.21$. Eccentricity too had a significant effect on the ratings (55.7, 50.7, and 48.3, for 0, ±60, ±120 eccentricities, respectively), $F(2, 46) = 8.03$, $p = 0.001$, $\eta^2_p = 0.26$. A repeated contrast test showed that this was caused by the higher ratings for the 0-eccentricity stimuli in comparison to the ±60-eccentricity stimuli. The interaction between eccentricity and identicalness was significant, $F(2, 46) = 7.81$, $p = 0.001$, $\eta^2_p = 0.25$, which was due to the fact that the effect of eccentricity was stronger for the identical-rectangles stimuli. To follow up on the interaction, we ran a separate ANOVA on the dissimilar-rectangles stimuli. The effect of eccentricity was marginally significant, $F(2, 46) = 2.67$, $p = 0.080$, $\eta^2_p = 0.10$. Contrasts showed that ratings became lower from the ±60 to the ±120-eccentricity (50.2 vs. 47.5), $F(1, 23) = 4.31$, $p = 0.049$, $\eta^2_p = 0.16$.

In the Balance task, the effect of identicalness was quite significant, with identical-rectangles stimuli being rated higher than dissimilar-rectangles stimuli (70.0 vs. 45.0), $F(1, 23) = 21.75$, $p < 0.001$, $\eta^2_p = 0.49$. The effect of eccentricity was highly significant too (63.6, 47.6, 37.4), $F(2, 46) = 29.29$, $p < 0.001$, $\eta^2_p = 0.56$. A repeated contrast test confirmed that both the step from 0 to ±60 and the step from ±60 to ±120 eccentricity partook in lowering the ratings. The interaction between eccentricity and identicalness was significant, $F(2, 46) = 13.98$, $p < 0.001$, $\eta^2_p = 0.38$, and the contrasts showed that this corresponded to the fact that the 0 to ±60 effect of eccentricity was larger in the identical-rectangles stimuli.

In the Weight task, we found higher ratings for identical-rectangles stimuli than for dissimilar-rectangles stimuli (57.9 vs. 47.6), $F(1, 23) = 21.01$, $p < 0.001$, $\eta^2_p = 0.48$. The effect of eccentricity was extremely large (76.7, 50.7, 31.1), $F(2, 46) = 223.65$, $p < 0.001$, $\eta^2_p = 0.91$, and a contrast test showed that both eccentricity steps contributed to lower the ratings. The interaction between eccentricity and identicalness, $F(2, 46) = 55.23$, $p < 0.001$, $\eta^2_p = 0.71$, was due to the fact that the higher ratings for identical-rectangles stimuli were observed only in the 0-eccentricity condition. In the two other conditions, dissimilar-rectangles stimuli received higher ratings.

The comparison between identical-rectangles stimuli and dissimilar-rectangles stimuli at the 0-eccentricity level (the two leftward bars in Figure 6, left) is especially interesting as it is the difference between symmetry and equilibrium. In the three tasks, this difference was highly significant, Aesthetics task: 76.1 vs. 52.7, $F(1, 23) = 15.42$, $p = 0.001$, $\eta^2_p = 0.40$; Balance task: 95.0 vs. 59.1, $F(1, 23) = 51.82$, $p < 0.001$, $\eta^2_p = 0.69$; Weight task: 99.2 vs. 73.4, $F(1, 23) = 131.40$, $p < 0.001$, $\eta^2_p = 0.85$. That is, the symmetric stimuli received a better aesthetics rating, are considered more balanced, and were judged to have the weight more centered than the equilibrated but asymmetric stimuli.

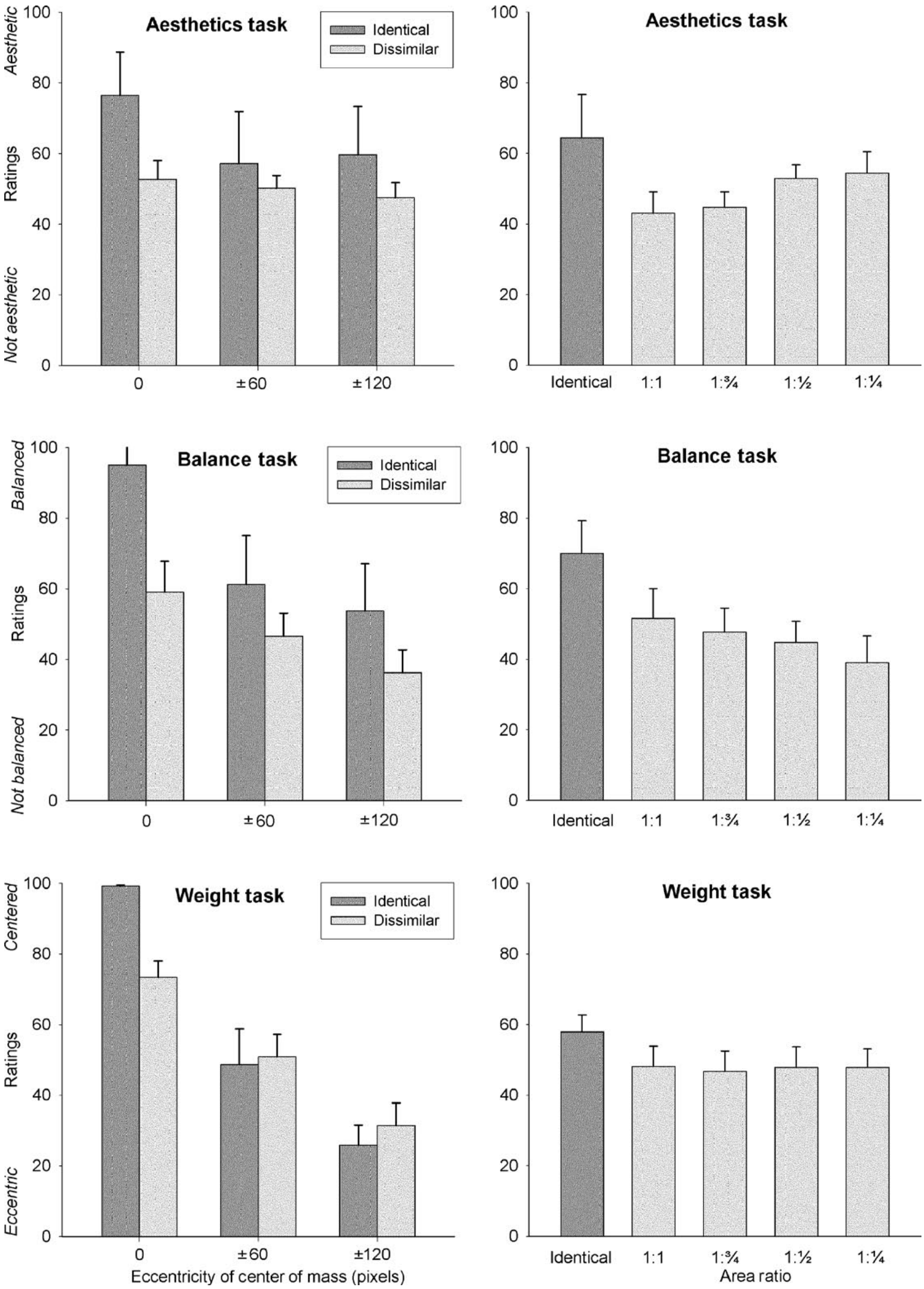

Figure 6. Left (top to bottom): the ratings in the three tasks plotted for the effect of the center of mass eccentricity. The dark gray columns give the values for the identical-rectangles stimuli; the light gray columns give the values for the dissimilar-rectangles stimuli. Right (top to bottom): the ratings in the three tasks plotted for the effect of ratio. The dark gray columns stand for the identical-rectangles stimuli (with 1:1 ratio); the light gray columns give the values for the dissimilar-rectangles stimuli in the four ratio situations.

3.2.3 The effect of area ratio

In Figure 6 (right), we see the results concerning the area ratio of the two rectangles composing the stimuli. The analyses were made exclusively on the dissimilar-rectangles stimuli, as the identical-rectangles stimuli by definition all have a 1:1 area ratio. After checking that there were no significant differences between the mirror ratios, we pooled across mirror ratios. In the Aesthetics task, we found a significant effect of area ratio, $F(3, 69) = 7.71$, $p < 0.001$, $\eta^2_p = 0.25$, and the repeated contrasts showed that this effect was entirely attributable to the difference between stimuli with no difference or with a small difference in area (ratios 1:1 and 1:¾; ratings: 44.7 and 43.0) relative to those with a large difference in area (ratios 1:½ and 1:¼; ratings: 52.9 and 54.5). That is, contrary to our hypothesis, stimuli composed of rectangles having quite different areas gathered higher ratings than those composed of rectangles having similar areas. The Balance task also exhibited an effect of ratio, $F(3, 69) = 4.87$, $p = 0.004$, $\eta^2_p = 0.18$, but the pattern of results was different than in the previous task. The balance ratings became increasingly lower with an increasing difference in area (51.6, 47.8, 44.7, 39.0). The Weight task produced no differences according to area ratio, $F(3, 69) < 1$, $p = 0.564$, $\eta^2_p < 0.10$ (48.1, 46.7, 47.8, 47.9, with increasing difference in area).

3.2.4 The effect of inter-rectangle gap

Concerning the size of the gap between the two rectangles, we did not hypothesize an interaction between dissimilar-rectangles stimuli and identical-rectangles stimuli, so the results for both types of stimuli are pooled. Neither the Aesthetics task, $F(3, 69) = 2.70$, $p = 0.052$, $\eta^2_p = 0.10$ (48.0, 51.4, 52.3, and 51.9, for the 10, 20, 30, and 50 pixel gaps, respectively), nor the Weight task, $F(3, 69) = 1.31$, $p = 0.280$, $\eta^2_p < 0.10$ (48.7, 47.8, 48.4, and 49.9), revealed an effect of gap. In the Balance task, we observed an effect of gap, $F(3, 69) = 3.94$, $p = 0.012$, $\eta^2_p = 0.15$, that is, contrasts showed that stimuli with the 50 pixel gap had higher ratings (52.9) than those with 10, 20, or 30 pixel gaps (46.0, 45.9, 47.7).

3.3 Discussion of Experiment 2

One effect is similar in the three ratings and replicates what we observed in Experiment 1. The difference in ratings between the symmetric stimuli and the equilibrated stimuli is highly significant. That is, equilibrated stimuli are rated less aesthetic, less balanced, and the weight is perceived as less centered than in the symmetric stimuli. This effect is strongest in the Weight task, although the two conditions should have received the same ratings, as in both cases the center of mass is in the center of the composition. However, the average for the Aesthetics task hides individual differences. Four participants, among which one had expressed artistic interest, rated the equilibrated composition much higher than the completely symmetric compositions, though they rated the symmetric compositions as having the weight more centered.

The correlations within the tasks are low, which means that between two mirror-symmetric versions of the same stimulus, ratings differ a lot. However, this is not too surprising as the shapes of the stimuli varied randomly. So this lack of reliability indicates that the exact shape matters. As the between-task correlations are even lower, we felt justified to look for differences between tasks.

A clear difference was that in the Aesthetics and Balance tasks, the identical-rectangles stimuli were rated highest whatever the position of the center of mass, but in the Weight task, this advantage disappeared when the center of mass was eccentric. Probably this was due to the fact that with identical-rectangles stimuli, eccentricity is most visible, which lowers the ratings in accordance with the task. The position of the center of mass had a general effect on the balance ratings and on the weight ratings. On the aesthetics ratings, the effect was weaker. Scores decreased for the identical-rectangles stimuli with a slight eccentricity only and decreased for the dissimilar-rectangles stimuli with a strong eccentricity only. The ratio of the rectangle areas also influenced the ratings differentially. We found that a large difference between the areas of the two rectangles influenced the perceived aesthetic value positively, while it influenced the perceived balance value negatively. That is, when the two rectangles have quite different areas, the stimulus is considered more aesthetic, but less balanced. However, not all participants' results followed this pattern. The weight ratings were not influenced by the ratio of the rectangle areas.

The gap between the two rectangles making up the stimuli had almost no influence. In the Balance task only, the largest gap produced slightly higher ratings, which might be explained by the fact that, in this task, the instructions were to rate whether the rectangles making up the composition were "well balanced, well distributed" which can be understood as "well spread out." And, indeed, a large gap spreads out the rectangles more than does a narrow gap.

4 General discussion

In two experiments we wanted to evaluate whether equilibrated compositions are considered as aesthetic as symmetric compositions. We also wanted to find out whether the focus of the rating was important: Do aesthetics ratings correspond to balance ratings and do balance ratings correspond to physical equilibrium ratings? This could be supposed as some authors, for instance Arnheim (1943, 1974) and Locher and colleagues (1996) declared that observers find aesthetic what is balanced, and balance, according to their statements, is due to forces that compensate. These statements would have been confirmed, if the equilibrated but asymmetric stimuli had obtained aesthetics ratings which were as high as those of the symmetric stimuli, or at least, if the equilibrated but asymmetric stimuli had obtained higher aesthetics ratings than the unequilibrated stimuli. In our experiments, this was not the case. Equilibrated but asymmetric stimuli were rated much lower than symmetric stimuli and not rated higher than the unequilibrated stimuli. So far our results are in accordance with McManus and colleagues' (2011) rejection of Arnheim's (1974) claim, and also in accordance with the study of Leyssen and colleagues (2012). That is, equilibrium is not a stimulus characteristic that guarantees high aesthetics ratings. However, in both experiments we found an effect of the strength of eccentricity of the center of mass. As stimuli representing strong disequilibrium received somewhat lower aesthetics ratings than those representing slight disequilibrium, we wondered why we failed to obtain highest ratings for the totally equilibrated stimuli. For the stimuli in Experiment 1, we think that the response is in the amount of discrepancy from symmetry, as the equilibrated stimuli were more asymmetric than those that were unequilibrated. In the equilibrated stimuli, the two lateral rectangles differed from each other as well in height as in width, whereas the unequilibrated stimuli differed only in one dimension, either in height or in width. Also, within the unequilibrated stimuli, except for the *stairs* shaped stimuli, height differences between the lateral rectangles lowered the ratings more than width differences between the lateral rectangles, probably because, with our stimuli, the height differences were detected more easily than the width differences. In the same vein, the lower ratings for compositions with strong disequilibrium in comparison to those with slight disequilibrium can be explained by the fact that the first ones also had a stronger asymmetry, as the differences in height or in width were larger. Thus, in Experiment 1, we have the impression that it is not equilibrium that is sought for, but rather the least perceived deviation from symmetry.

The picture is somewhat different in Experiment 2, where equilibrium is varied independently from symmetry. In the identical-rectangles stimuli, we found a strong drop in the aesthetics ratings when the center of mass is not centered. This drop confirms that participants did not neglect the overall limit of the composition, as they react so strongly when the identical-rectangles stimuli become eccentric. But it is the dissimilar-rectangles stimuli that allow us to judge the effect of equilibrium when there is no symmetry. For these stimuli, we observed a slight decrease in ratings from slight to strong eccentricity of center of mass. These results tell us that there is some effect of equilibrium. However, this effect is weaker than what we would expect according to Arnheim's (1943, 1974) and Locher and colleagues' (1996) writings.

Observers could have failed to give higher aesthetics ratings to equilibrated stimuli because they could have been unable to judge whether the compositions are equilibrated. Previous studies showed that observers are able to judge equilibrium, though not very accurately and not without biases (Barnett-Cowan, Fleming, Singh, & Bülthoff, 2011; Samuel & Kerzel, 2011). In the present study, we ran the Weight task to investigate whether observers were able to judge the equilibrium state of our stimuli. In both experiments, the clearest result was that totally equilibrated stimuli were considered to have the center of mass less in the center than symmetric stimuli. Therefore, we can say that the failure to rate equilibrated stimuli as very aesthetic is partially attributable to the fact that they are not perceived as equilibrated. But are equilibrated stimuli at least perceived as having the weight more centered than the unequilibrated stimuli? In Experiment 1, the evidence is unclear, as the higher ratings in the Weight tasks for the equilibrated stimuli in comparison to the unequilibrated stimuli was attributable exclusively to the low ratings given to the height-modified stimuli; width-modified stimuli were not rated lower than the equilibrated stimuli. But in Experiment 2, there was no doubt that the weight ratings became much lower with increasing center of mass eccentricity, which means that observers are able to judge equilibrium. None of the participants was an exception to that. So the aesthetics ratings should have followed the same decline with increasing eccentricity, if the perceived center of mass eccentricity was the standard for the aesthetics ratings. This was not the case for most participants.

Given that the state of equilibrium had not more than a slight or inconsistent influence on the aesthetics ratings, what were the other influences on these ratings? In Experiment 1, the effects were not those we expected. Looking at the effect of shape, we thought that the *stairs* stimuli would receive low ratings, as they are so obviously asymmetric. But in the aesthetics ratings, the participants' preference for this shape is quite noticeable. We suppose that participants appreciated the progressive decrease or increase of height. This form of regularity might be nearly as pleasing as the regularity of symmetry. Maybe the *stairs* shape enhances the symmetry relative to the *horizontal* axis (which all stimuli had, but which might be more perceptible in some) and we know from Bauerly and Liu (2006) that symmetry with a horizontal axis is rated aesthetic too. Concerning the two other shapes, we expected *mid-high* shaped stimuli to receive better ratings than the *mid-low* shaped stimuli because the former seemed to offer better Gestalt properties due to their overall rounded shape, their compactness, therefore their better *closure* than the latter ones (Hartmann, 1935; Reuchlin, 1998). However, the aesthetics ratings did not differ between these two shapes. Effects related to shape were also visible in Experiment 2. Surprisingly, stimuli made up of rectangles of very different areas were aesthetically more appreciated than when the two rectangles had the same or nearly the same areas. We conclude that when there is no hint of symmetry, because the two rectangles are dissimilar, then observers prefer when the two rectangles are markedly different in comparison to slightly different.

We were especially interested to observe whether variables other than the position of the center of mass influenced balance and weight ratings in a different way than they influenced the aesthetics ratings. In the Balance task, the effect of shape was somewhat different than in the Aesthetics task. As in the Aesthetics task, the *stairs* shape was rated higher, but as a supplementary effect to the one found in the Aesthetics task, *mid-low* stimuli were rated higher (more balanced) than *mid-high* stimuli. In the Weight task, the higher ratings for the *stairs* shape disappeared completely, but the advantage of *mid-low* over *mid-high* stimuli was carried on from the Balance task. That is, the effect of shape on the balance ratings was intermediate between the one on aesthetics ratings and the one on weight ratings. The effects due to area ratio differentiated the three tasks more drastically: In the aesthetics ratings a large difference in area improved the ratings, whereas in the balance ratings a large difference in area lowered the ratings. In the Weight task, there was no effect of area ratio.

Finally, we would like to point out a potential caveat in the present study. For most participants, we found low or absent correlations from one task to another, which confirms that the tasks elicited quite differing ratings. However, the within-task correlations of many participants were also quite low, which indicates that the ratings differed where they were expected to be similar. In both experiments, this unreliability was strongest in the Aesthetics task. Furthermore, we observed that patterns of results differed across participants. For example, a few participants aesthetically preferred the asymmetric over the symmetric compositions, which contrasted with the overall results.

5 Conclusions

Our findings highlight that aesthetics ratings should not be considered equivalent to balance ratings, and balance ratings are not the same as physical equilibrium ratings. However, for a few participants, the three ratings followed similar trends. Furthermore, our results emphasize that any statement about the effect of equilibrium must be made when equilibrium is bare of symmetry, as the effect of symmetry or approximate symmetry overshadows all other effects.

Acknowledgments. The experiments were made using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience. Many thanks to William Larson for reading and correcting the manuscript.

References

Arnheim, R. (1943). Gestalt and art. *The Journal of Aesthetics and Art Criticism, 2*(8), 71–75. doi:10.2307/425947

Arnheim, R. (1974). *Art and visual perception*. Berkeley and Los Angeles: University of California Press.

Barlow, H. B., & Reeves, B. C. (1979). Versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19*(7), 783–793. doi:10.1016/0042-6989(79)90154-8

Barnett-Cowan, M., Fleming, R. W., Singh, M., & Bülthoff, H. H. (2011). Perceived object stability depends on multisensory estimates of gravity. *PLoS One, 6*(4), e19289. doi:10.1371/journal.pone.0019289

Bauerly, M., & Liu, Y. (2006). Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human–Computer Studies, 64*(8), 670–682. doi:10.1016/j.ijhcs.2006.01.002

Campell, I. G. (1941). Factors which work toward unity or coherence in visual design. *Journal of Experimental Psychology, 28*(2), 145–162. doi:10.1037/h0056758

Frances, R. (1967). L'esthétique expérimentale: Rapport de l'esthétique avec la psychologie. *Bulletin de Psychologie, 20*(8–9), 575–584.

French, J. E. (1956). Children's preferences for abstract designs of varied structural organization. *The Elementary School Journal, 56*(5), 202–209. doi:10.1086/459445

Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception, 2*, 508–527. doi:10.1068/i0472aap

Gombrich, E. H. (1982). *The image and the eye: Further studies in the psychology of pictorial representation*. Oxford: Phaidon.

Gottlieb, C. (1958). Movement in painting. *The Journal of Aesthetics and Art Criticism, 17*(1), 22–33. doi:10.2307/428007

Hartmann, G. W. (1935). *Gestalt psychology: A survey of facts and principles*. New York: Ronald Press.

Leyssen, M. H. R., Linsen, S., Sammartino, J., & Palmer, S. E. (2012). Aesthetic preference for spatial composition in multi-object pictures. *i-Perception, 3*, 25–49. doi:10.1068/i0458aap

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. doi:10.2190/EKMD-YMN5-NJUG-34BK

Locher, P., Gray, S., & Nodine, C. (1996). The structural framework of pictorial balance. *Perception, 25*(12), 1419–1436. doi:10.1068/p251419

Locher, P., & Nagy, Y. (1996). Vision spontaneously establishes the percept of pictorial balance. *Empirical Studies of the Arts, 14*(1), 17–31. doi:10.2190/X8U3-CTQ6-A7J1-8JQ8

Locher, P., 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. doi:10.1016/S0001-6918(98)00008-0

Lund, F. H., & Anastasi, A. (1928). An interpretation of aesthetic experience. *The American Journal of Psychology, 40*(3), 434–448. doi:10.2307/1414460

McManus, I. C. (2005). Symmetry and asymmetry in aesthetics and the arts. *European Review, 13*(2), 157–180. doi:10.1017/S1062798705000736

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

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*, 615–647. doi:10.1068/i0445aap

Reuchlin, M. (1998). *Psychologie* (13th ed.). Paris: Presses Universitaires de France.

Samuel, F., & Kerzel, D. (2011). Is this object balanced or unbalanced? Judgments are on the safe side. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 529–538. doi:10.1037/a0018732

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences, 1*(9), 346–352. doi:10.1016/S1364-6613(97)01105-4

Wenderoth, P. (1994). The salience of vertical symmetry. *Perception, 23*(2), 221–236. doi:10.1068/p230221

Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts, 23*(2), 165–180. doi:10.2190/B1LR-MVF3-F36X-XR64

Appendix A: Example of calculation of the center of mass position

An example of the calculation of the center of mass position is given in Figure A1.

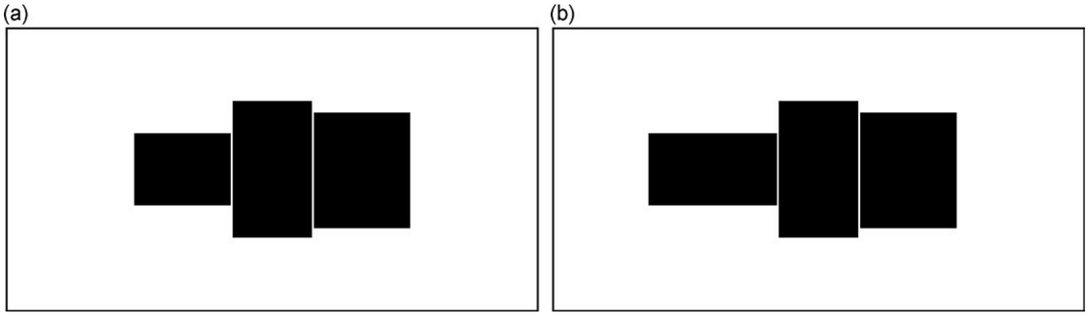

Figure A1. Example of calculation of the center of mass position. (a) Modified height stimulus (length indications in pixels; mass = area; indications in square pixels): Left R: W × H: 170 × 128 → mass: 21,760; x-location: −(70 + 85 + 1) = −156. Mid. R: W × H: 140 × 240 → mass: 33,600; x-location: 0. Right R: W × H: 170 × 204 → mass: 34,680; x-location: 70 + 85 + 1 = +156. Center of mass = [(−156 × 21,760) + (0 × 33,600) + (156 × 34,680)]/(21,760 + 33,600 + 34,680) ≈ 22. That is, the center of mass is 22 pixels to the right of the center of the middle square. (b) Equilibrium stimulus (length indications in pixels; mass = area; indications in square pixels): Left R: W × H: 228 × 128 → mass: 29,184; x-location: −(70 + 114 + 1) = −185. Mid. R: W × H: 140 × 240 → mass: 33,600; x-location: 0. Right R: W × H: 170 × 204 → mass: 34,680; x-location: 70 + 85 + 1 = +156. Center of mass = [(−185 × 29,185) + (0 × 33,600) + (156 × 34,680)]/(29,184 + 33,600 + 34,680) ≈ 0. That is, the center of mass is in the center of the middle square.

Françoise Samuel worked as an artist painter for many years. Then she studied psychology at the Universities of Lausanne and Geneva. Her research interests, in accordance with her first profession, went to visual perception and visual aesthetics. Her doctoral dissertation, which she did under the supervision of Professor Dirk Kerzel, concerned the visual perception of the equilibrium state of objects. Now she works as a postdoctoral assistant at the University of Geneva.

a Pion publication