

Arnheim's Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images

I C McManus

Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, UK; e-mail: i.mcmanus@ucl.ac.uk

Katharina Stöver

Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, UK; e-mail: k.stoever@ucl.ac.uk

Do Kim

Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, UK; e-mail: ucjtdyk@ucl.ac.uk

Received 12 May 2011, in revised form 12 September 2011; published online 19 October 2011

Abstract. In *Art and Visual Perception*, Rudolf Arnheim, following on from Denman Ross's *A Theory of Pure Design*, proposed a Gestalt theory of visual composition. The current paper assesses a physicalist interpretation of Arnheim's theory, calculating an image's centre of mass (CoM). Three types of data are used: a large, representative collection of art photographs of recognised quality; croppings by experts and non-experts of photographs; and Ross and Arnheim's procedure of placing a frame around objects such as Arnheim's two black disks. Compared with control images, the CoM of art photographs was closer to an axis (horizontal, vertical, or diagonal), as was the case for photographic croppings. However, stronger, within-image, paired comparison studies, comparing art photographs with the CoM moved on or off an axis (the 'gamma-ramp study'), or comparing adjacent croppings on or off an axis (the 'spider-web study'), showed no support for the Arnheim–Ross theory. Finally, studies moving a frame around two disks, of different size, greyness, or background, did not support Arnheim's Gestalt theory. Although the detailed results did not support the Arnheim–Ross theory, several significant results were found which clearly require explanation by any adequate theory of the aesthetics of visual composition.

Keywords: aesthetics, photography, Gestalt theory, balance, composition.

1 Introduction

Composition, the organisation of objects and forms in relation to one another, is central to the production of works of art, and hence to aesthetics, be it in the visual arts, music, literature, or other art forms. Gustav Theodor Fechner, in his *Vorschule der Aesthetik* (1876), pointed out that

[forms] are never used in isolation, but always with neighbouring shapes and proportions ... [together these result in what] I called the combinatorial determining influence (Fechner 1997, p. 118, our emphasis).

Art works typically have boundaries, which in the case of paintings, drawings, or photographs are represented by a frame, which most often is rectangular, and may itself be a work of art (Gombrich 1979) or, in modern works, may be minimal, being indicated primarily by the image being in a different plane to the wall on which it is hung. Given a frame, then the problem of composition becomes the arrangement of objects and of patterns of light and shade within that boundary. Working artists typically know what is meant by good composition, although defining the term, or dissecting its components, is far from straightforward. Even less is known about the necessity of composition, for as the theoretician of photography Victor Burgin has put it:

We know very well what good composition is—art schools know how to teach it—but not why it is; 'scientific' accounts of pictorial composition tend merely to reiterate what it is under a variety of differing descriptions (eg, those of Gestalt psychology) (Burgin 1982, p. 150, emphasis in original).

The reference to Gestalt psychology is interesting, as there continues to be an interest in Gestalt theory among artists (Behrens 1998; 2002), and the scientific literature itself contains few other accounts of the nature of composition except those of Gestalt theorists.

Gestalt psychology emphasised that some forms, such as say a circle, are 'good', in the sense that they are simpler or are easier to recognise or describe. Descriptions of the Gestalt principles of good organisation (closure, similarity, proximity, symmetry, common fate, etc) can be found in many introductory and advanced textbooks of vision (eg, Metzger 2006; Gordon 2004). The intellectual origins of Gestalt theory can be found in the field theories of electromagnetism, which were described by physicists such as James Clark Maxwell (Ash 1998). To a physicist a circle or sphere is formed, as, say, in a soap bubble, when a system finds an equilibrium at a minimum energy level. Using a similar argument, the Gestalt theorists, Kohler in particular, suggested that the goodness of 'good forms' (gute Gestalten) occurs because the perception of the object, such as a circle on a page, results in a similar pattern of activity on the surface of the brain, a so-called isomorphic representation of the circle, which also has a minimum energy. Such ideas, which Verstegen (2007, p. 37) has referred to as 'the notorious Gestalt brain model', are, of course, completely incompatible with modern understanding of the neuroscience of vision so that although the Gestalt principles are in every textbook, isomorphism is now nothing but a historical curiosity. That has, however, left a difficult theoretical gap, for although the Gestalt principles seem to be sound as descriptions of perceptual phenomena, there is no strong theoretical underpinning for their origin. On the one hand, it ought to be possible, as Verstegen (2007, p. 39) suggests, to keep the field metaphor and jettison the specific mechanisms. On the other hand, that is the precise opposite of Cupchik, who says, 'Concepts such as fields and forces possess surplus meanings grounded in other disciplines ... and are best treated in a metaphorical and reflective manner so they don't become reified' (Cupchik 2007, p. 23).

Within psychological aesthetics one of the most important of writers has been Rudolf Arnheim (1904–2007), much of whose theorising originates within Gestalt psychology (Cupchik 2007; Verstegen 2007). In *Art and Visual Perception* (Arnheim 1954; 1974)⁽¹⁾ and *The Power of the Center* (1982), Arnheim emphasised an holistic interpretation of an entire image or space, in which there are interacting, balancing sets of forces. Despite the suggestion of Cupchik (2007) that the use of terms such as 'forces' should be considered as metaphorical, and he was writing specifically within the context of a critical reflection on Arnheim's theory, a close reading of Arnheim suggests that he did indeed want his theory to be taken in a fairly literal sense (and indeed, when talking about 'perceptual forces', Arnheim is very clear, since after speculating whether these forces are, 'merely figures of speech, or are they real', he continues, 'they are assumed to be real in both realms of existence—that is, as both psychological and physical forces' (Arnheim 1974, p. 16).

In interpreting Arnheim, and the simple designs that he considers in *Art and Visual Perception*, it is important also to realise that much of Arnheim's theorising derives from Denman Ross's *A Theory of Pure Design* (1907), which Arnheim cites, and which takes a much more explicitly physicalist approach. Denman Ross's book has been much neglected in the psychological literature, to the extent that in Web of Science we found only 11 citations, only 3 of which were from psychology journals (Daniels 1933; Jacobson 1933; Jasper 1933). Nevertheless, the book is of great interest, not least because Denman W Ross (1853–1935)

⁽¹⁾All quotations are taken from the 1974 edition.

was a polymath in the humanities and a lecturer in the Department of Fine Arts at Harvard (Hopkinson 1937), the university in which Arnheim was based when he prepared the revised edition of *Art and Visual Perception*. Although not a scientist, Ross does say in the preface to *A Theory of Pure Design* that although, 'Art is regarded as the one activity of man which has no scientific basis', it is nevertheless his, 'purpose ... to show how, in the practice of Art, as in all other practices, we ... follow certain principles' (1907, p. v). Elsewhere Ross was to talk of, 'Design as a science' (Ross 1901), and he clearly had a desire for '“measurable quantities and qualities” of art' (Johnson 1995, p. 59), so that he was aiming for, 'a scientific basis for design, [which could be] a verification or correction of visual feeling' (Ross 1901, pp. 371–72). In the present paper we will firstly consider how the Arnheim–Ross theory of visual composition can be formalised so that it makes specific empirical predictions, and then test the theory directly by means of several experiments, mostly involving photographic images.

Chapter 1 of *Art and Visual Perception* is entitled 'Balance' and is devoted to the question of composition in visual works of art. The chapter begins with a reductionist approach that is reminiscent of the approach of a physicist, considering the relationship between a single black disk placed on a white background within a square frame. When the disk is off-centre, 'There is something restless about it. ... The disk's relations to the edges of the square [frame] are a ... play of attraction and repulsion' (Arnheim 1974, p. 11). The frame has an influence upon the disk, and indeed even an empty square frame,

is empty and not empty at the same time. Its center is part of a complex hidden structure, which we can explore by means of the disk, much as we can use iron filings to explore the lines of force in a magnetic field (p. 12, our emphasis).

Arnheim's approach is undoubtedly derived from Ross's *A Theory of Pure Design* (1907), which Arnheim cites on page 19. Ross discusses the problem of achieving a visual balance for a set of small disks arranged asymmetrically.

This is best done by means of a symmetrical inclosure or frame. In ascertaining just where the centre is ... we depend upon visual sensitiveness or visual feeling, guided by an understanding of the principle of balance: that equal attractions, tensions or pulls, balance at equal distances from a given center, that unequal attractions balance at distances inversely proportional to them. Given certain attractions, to find the center, we weigh the attractions together in the field of vision and observe the position of the center (Arnheim 1974, p. 23, our emphasis).

It is hopefully clear that what Ross is describing is not mere metaphor, but is intended to be taken literally, in a sense that a physicist would recognise, and that in principle the balance can be computed as if it were a problem in physics. Ross is also sensitive to the subtleties that should occur when the objects are not mere black disks but 'vary in their tones, measures, and shapes, and where there are qualities as well as quantities to be considered' (p. 23), although he acknowledges that in such cases, 'calculations and reasoning becomes difficult if not impossible, and we have to depend upon visual sensitiveness' (p. 23), the visual system presumably carrying out such calculations. In a direct analogy with the cropping methodology to be described later, a composition which is in balance can be found by moving a frame around the set of dots or objects:

Move the frame up or down, right or left, and the center of the frame and center of the attractions will no longer coincide, and the balance will be lost. We might say of this arrangement that it is a Harmony of Positions due to the coincidence of two centers, the center of the attractions and the center of the framing (p. 24).

This methodology is mentioned by Arnheim, who after describing Ross's work, says, 'the center of the frame coincides with the weight center of the pattern' (p. 19, our emphasis).

The only important difference between Arnheim and Ross is that while Ross seems to require the centre of the attractions to be at the physical centre of the frame, Arnheim

suggests that a picture can be balanced if the centre of the attractions is placed on any of the major axes of an image, the horizontal or vertical axes, or of the two diagonals—and Arnheim's Figure 3 (p. 13) shows that 'a disk is influenced not only by the boundaries and the centre of the square [the frame], but also by the cross-shaped framework of the central vertical and horizontal axes and by the diagonals' (p. 13), which are 'comparatively restful positions' (p. 14). The psychological literature on balance has implicitly, and often explicitly, adopted physicalist ideas of balance [eg, in earlier work of one of us (McManus et al 1985; McManus and Kitson 1995), as well as that of Locher et al (1996)]. Sometimes this involves simple synthetic images, often only consisting of black and white, rather than the multiple grey levels of photographs, as, for instance, in Locher et al (2001), Locher et al (1998), and Wilson and Chatterjee (2005), or in images drawn from Japanese calligraphy (Gershoni and Hochstein 2011), in all of which studies a balance score is calculated.

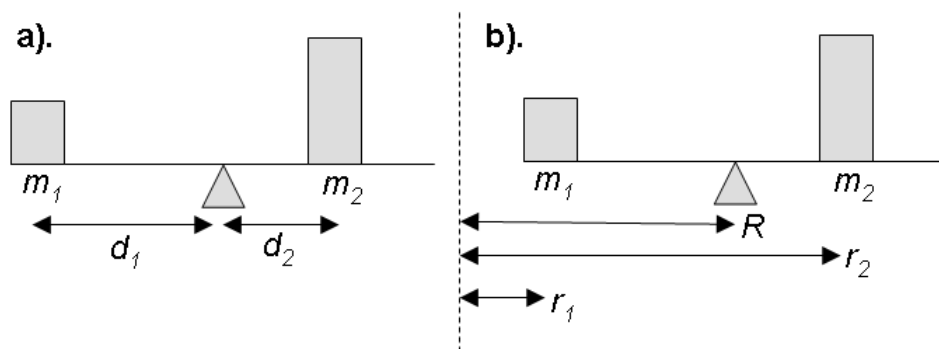


Figure 1. The calculation of the centre of mass of two physical objects arranged on a beam, which has a fulcrum, indicated by the triangle, at the centre. (a) The objects have mass m_1 and m_2 and are at horizontal distances d_1 and d_2 from the fulcrum. The beam is balanced if $m_1 \cdot d_1 = m_2 \cdot d_2$. (b) More generally, if the positions of the two masses and the fulcrum are measured as distances, r_1 , r_2 , and R from an arbitrary horizontal position shown by the vertical dashed line, then the beam is balanced when $R = (m_1 \cdot r_1 + m_2 \cdot r_2) / (m_1 + m_2)$ (see text), with the fulcrum being at the centre of mass.

Ross's account, and by implication Arnheim's, is physicalist, as when Ross talks of how, 'unequal attractions balance at distances inversely proportional to them'. Were Arnheim not intending to be physicalist, then, having cited Ross, he would surely have felt obliged to clarify that his own interpretation was only metaphorical. For most of the rest of this paper we will interpret the theory physically, using the standard ideas of elementary physics on identification of a centre of gravity, or more generally the centre of mass (CoM), of a set of objects. Figure 1a shows the situation of a simple balance, in which two objects of different mass, m_1 and m_2 , are located at different horizontal distances, d_1 and d_2 , from the fulcrum of the balance. The system is in static equilibrium, and tilts neither to left nor right if $m_1 \cdot d_1 = m_2 \cdot d_2$. In practice, a more useful calculation is to find the position of the fulcrum in relation to some arbitrary horizontal position, so that as in Figure 1b, the fulcrum and the two masses are at distances R , r_1 , and r_2 from the arbitrary location. The fulcrum is then located at a distance R from the arbitrary location, where $R = (m_1 \cdot r_1 + m_2 \cdot r_2) / (m_1 + m_2)$. That formula then generalises to n objects, using equation 1. Formula 1 can locate the horizontal CoM of any number of objects (and a directly equivalent formula can locate the CoM in the vertical or any other direction). In particular, as Figure 2 shows, the elements can be the pixels of a photographic image in which the 'mass' or 'weight' of each pixel consists of the inverse of its grey level (eg, white is regarded as 'light', and scored as 0, and black is regarded as 'heavy' and scored as 255, with other grey levels in between). It is therefore straightforward, for any grey-level image, to calculate the centre of mass horizontally and vertically. An example

analysis of a monochrome image is shown in [Figure 3](#). Superimposed on the image are several sets of lines. In particular, the blue line, g, running from left to right shows the average lightness (up = light, down = dark) of each of the 1,024 columns of pixels, and the yellow line, h, running from top to bottom shows the average lightness (left = dark, right = light) of each the 768 horizontal rows of pixels. Using the formula provided above, those values can be used to locate the horizontal and vertical positions of the CoM, which is shown as the solid white square, i, which is .477 of picture width from the left of the image, and .455 of picture height from the bottom of the picture. The Arnheim–Ross model suggests that the CoM should be located on one of the main axes of the image, and the horizontal and vertical midlines, and the two diagonals are shown as solid red lines (c, d, e, f). The CoM in this case is not on any of the four axes specified by Arnheim. Expressed in units of picture width/height, the CoM is .023 units from the horizontal midline, .045 from the vertical midline, .048 from the major diagonal, and .016 from the minor diagonal, the axis to which it is closest. In this paper we will define an image as 'on axis' when its CoM is within + 0.5% of standardised picture width/height of the vertical, horizontal, or diagonal axes. In the case of [Figure 2](#) it is therefore the case that the centre of mass does not meet that criterion, and although the closest axis is the minor diagonal, the image as a whole is 'off axis'.

$$|R| = \frac{\sum_{i=1}^n m_i \cdot r_i}{\sum_{i=1}^n m_i} \tag{1}$$

The present paper tests some detailed predictions resulting from this physicalist interpretation of the Arnheim–Ross model, described above, assessing the extent to which images that are aesthetically satisfactory or superior have their centres of mass on one of the major axes, as Arnheim and Ross suggest. The five studies are complex, and readers might find it helpful to have an overview of the structure of the paper and its broad conclusions at this stage. Study 1 investigates the location of the centre of mass in a large set of art photographs printed in *The Photography Book* (Jeffrey 2005), comparing their location with an internal set of control images generated by sampling from the art photographs, and an external set of control images consisting of randomly taken photographs. To anticipate the results, art photographs do indeed have their centres of mass closer to a major axis than do control images, which supports the Arnheim–Ross theory, although the test is relatively weak, since other processes may have resulted in those differences. Study 2 therefore provides a stronger test of the Arnheim–Ross theory, in which art photographs are manipulated, by means of altering gamma (in effect the contrast) of the images by different amounts in the horizontal and vertical directions, so that centres of mass are either moved to or from an axis. In a paired comparison design, however, there was no evidence that subjects preferred images in which the centre of mass was on an axis, providing no support for the Arnheim–Ross model.

Studies 3, 4, and 5 used a cropping paradigm, which we have described elsewhere (McManus et al 2011) in which the mouse of a computer is used to select a 512 x 384 region of a 1024 x 768 image displayed on a computer screen, in a direct analogue of the way in which a camera is used to select or crop a portion of the visual world when a photograph is taken. The cropping task is also a close simulation of the procedure that Denman Ross described: of moving a frame to determine 'just where the centre is ... depend[ing] upon visual sensitiveness or visual feeling'. Study 3 examined the photographic crops generated in the previous study (McManus et al 2011) and compared their locations with the loci, suggested by the physical theory described earlier, which should produce balance with the centre of mass lying on one of the major axes. As the Arnheim–Ross model predicts, the centres of

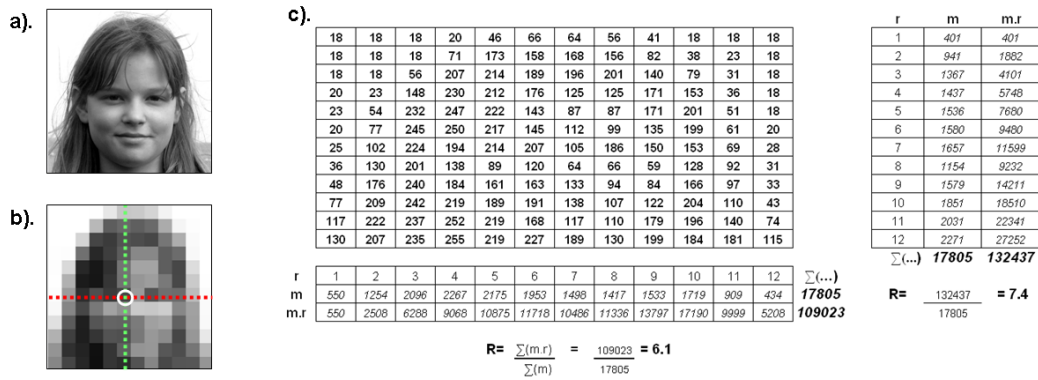


Figure 2. Example of calculation of the centre of mass of an image. (a) The original square image; (b) A 12 x 12 (pixelated) version of the image so that the calculations are simpler to see (although for the images proper in the study all pixels are used from the entire image); (c) The calculations. The 12 x 12 matrix shows the darkness or weight of each pixel (ie, 255 = black = heavy; 0 = white = light). The table at the bottom show r_i , the arbitrary horizontal distance of each pixel from the left-hand side, expressed as pixel number, the summed mass of each column of pixels, m_i , and the product of m_i and r_i ($m_i \cdot r_i$). At the end of each row are the sums of r_i and $m_i \cdot r_i$, which are used to calculate R, the position of the centre of mass (see main text for details), which is at 6.1 pixels from the left-hand side. The table at the right shows the equivalent calculations for the rows, with r_i arbitrarily being the vertical distance from the top row of pixels. The centre of mass for the image as a whole is 7.4 pixels from the top and 6.1 pixels from the left, and those positions are shown by the red and green dashed lines in (b), with the intersection, shown by the white circle, being the centre of mass of the image.

mass of cropped images were closer to major axes than were control images, although this is a relatively weak test. Study 4 extended the analysis using a stronger test in what we have called the 'spiders-web' experiment; by analogy with Study 2, pairs of images were derived from the same images cropped in experiment 2, but one member of each pair was on a major axis,⁽²⁾ and the other matched image was equidistant from adjacent major axes. In Study 4 there was no support for the Arnheim–Ross model. Study 5 also used the cropping paradigm, but instead of using real photographs, it instead used much simpler images based on the example described by Arnheim, of two black disks, around which subjects were asked to position a frame, as Ross had suggested. The results with two black disks were broadly as Arnheim would have predicted, although that is a weak test. However, a stronger test considers the situation when one of the disks is grey rather than black, is smaller than the other disk, or when there is a grey-level ramp, running either vertically or horizontally across the field. Although calculation in those cases suggests that croppings should be radically different from Arnheim's example of two identical black disks, our participants framed the images in precisely the same way as they framed two black disks, suggesting that the detailed Arnheim–Ross model is not supported.

2 Study 1: The centre of mass in art photographs

2.1 Study 1: Introduction

Any theory of aesthetics has to predict that works of art that are part of the recognised canon should be different in some way from those that are of lesser quality, having been produced

⁽²⁾A note on nomenclature. 'Horizontal' and 'vertical' are potentially very confusing. The axis running from top to bottom, is the vertical axis, which runs through the horizontal midline of the image, and if the image is balanced, the horizontal CoM is on the vertical axis. We will therefore refer to the vertical axis (which relates to the horizontal CoM and the horizontal midline) and the horizontal axis (which relates to the vertical CoM and the vertical midline).

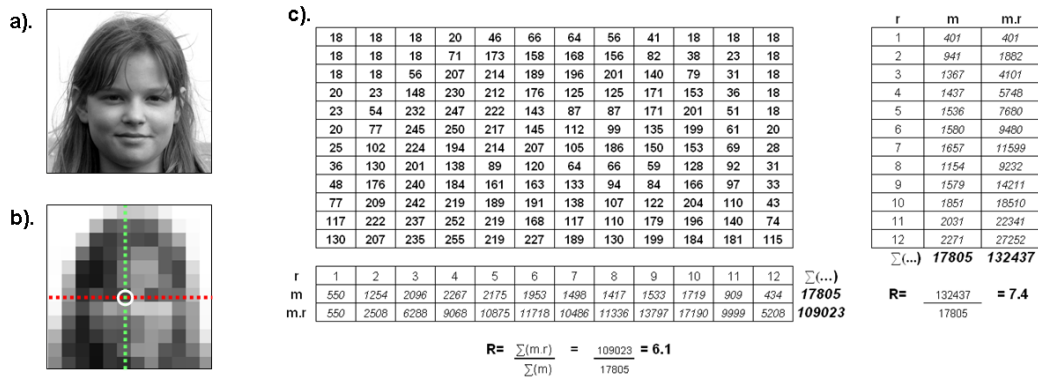


Figure 3. Illustration of the calculation of the centre of mass of a photograph, along with locations of centres of mass of internal control images. Main image (a) shows a monochrome photograph of a coot, along with (b) an enlarged detail from the centre of the analysed image. The vertical (c) and horizontal (d) red lines indicate the vertical axis and the horizontal axis of the image, and the red diagonal lines indicate the major diagonal axis (e) and the minor diagonal axis (f). The blue line (g) going from left to right indicates the lightness (up) or darkness (down) of each column of the image, and the yellow line (h) from top to bottom indicates the lightness (to right) or darkness (to left) of each row of the image. The centre of mass of the image overall is shown by the white square with a black surround (i), and the 289 (or 17 x 17) small coloured squares with black surrounds (eg, j, k, l, and m) correspond to the scaled centres of mass of the internal control images generated by systematically sampling all images of half the width and height of the complete image. In particular, j, k, l, and m correspond to the top left, top right, bottom left, and bottom right quarters of the image, respectively (although the square for j is overlaid with i). Original photograph is by one of the authors (ICM).

by individuals who are not artists or representing works that are degraded in some way or another. Although Arnheim in his later works avoided referring to 'the law of Prägnanz', nevertheless, as Verstegen has emphasised in a section entitled ' "Goodness" and beauty', 'the idea that "good" Gestalten placed one on the road to explaining esthetic phenomena was what so excited Arnheim in his early work' (Verstegen 2007, p. 12). It therefore seems a straightforward prediction from the Arnheim–Ross theory that accepted works of art of quality should show greater balance than those of lesser quality. That idea was tested in Study 1 by looking at a large sample of representative 'art photographs' (for want of a better term). Clearly, any such study requires a control group, and we describe two different types of controls, internal and external.

2.2 Study 1: Method

2.2.1 Art photographs.

These images were obtained from *The Photography Book* (Jeffrey 2005), a high quality, large format collection of 500 photographs, one each by 500 famous photographers, arranged alphabetically from A to Z, from Aarsman, Abbas, and Abbe to Zachmann, Zecchin, and Van der Zee. This corpus has the important advantage of being representative, and having only a single image from each photographer, we thereby avoid statistical problems of non-independence. Photographs were included in the present study if they were single, rectangular images, without obvious technical blemishes or post-production alteration (eg, writing, painting, and so on). Images were scanned on a professional quality scanner (Epson Express 1640 XL), set at 300 dots per inch, 24 bit colour, the unsharp masking filter set at medium, and the descreening filter in general mode. It should be noted that the aspect ratio of the art photographs was not the same for all images.

2.2.2 External controls.

A series of 'random photographs' was taken by one of the authors (ICM) using a Canon Ixus 82 IS digital camera, with image size 2048 x 1536 pixels, with standard colour settings, autofocus and exposure, and the display turned off. The camera

was mostly held horizontal (except when photographing the ground) while walking down streets or parks, sitting on buses or trains, in buildings, or other locations. An explicit attempt was made to avoid pointing the camera at objects (and hence framing them), and often the camera was held at the photographer's side so that he was not looking directly at the subject of the photograph. Likewise, the shutter was pressed, where possible, at regular time intervals, such as every 15 s, to avoid bias. Similarly, the optical zoom of the camera was also operated at random, to ensure images were as different as possible from one another. Overall, 595 photographs were available for use.

2.2.3 Image analysis. Special purpose software was written using Matlab 7.5 with the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Images in colour were converted to monochrome using the `decolorize` routine, which is also written in Matlab (Grundland and Dodgson 2007), and which has been shown psychophysically to give the most satisfactory conversion perceptually from colour to monochrome (Cadik 2009). The basic analysis of images followed that which was described earlier and is described below in more detail and in the caption to Figure 3. Because the aspect ratio of the art photographs differed, the location of the centre of mass was described in terms of the proportion of the image width or height (0 = left-hand side or bottom edge; 1 = right-hand side or top edge).

2.2.4 Internal controls. An ideal control for any image would consist of scenes adjacent to the image that the photographer might have chosen, but did not, since those images would have similar contrast, lightness, colour balance, Fourier descriptors, and so on as those in the original image, as well as similar content. Needless to say, such images are usually not available. A compromise is to use alternative images within the actual photograph, a process that can be conceptualised as examining sub-images that the photographer might have taken had they used a lens of twice the focal length; such sub-images are half the width and half the height (and hence a quarter of the area) of the original image. We looked at all possible positions of the sub-image on a grid within the original image, dividing the height and width of the original into 16 parts (ie, 17 possible positions for the left-hand edge of the sub-image, from the extreme left of the original image to the midline of the original, and likewise 17 possible vertical positions for the top of the sub-image, from the top of the original image to its vertical midline). The CoM of these $17 \times 17 = 289$ sub-images was calculated in the standard way and expressed as the proportion of the width and height of the sub-image. The CoMs, expressed as proportions of image size, can then be plotted on the original image (and for comparative purposes, this is in terms of the proportions of the original image)—ie, if the CoM of a sub-image is exactly at the centre of the sub-image, then in Figure 3 it is plotted exactly at the centre of the original image so that if the corners of the sub-image are at $[s, s]$ and $[s + .5, s + .5]$, and the CoM is at (a, b) , then the CoM is plotted at $[2(a - s), 2(b - s)]$.

2.3 Study 1: Results

Overall, 455 art photographs and 595 random control images were analysed.

2.3.1 Location of CoMs. Figure 4 shows the CoMs of the art photographs and the random photographs. The Arnheim–Ross model predicts that CoMs of art photographs should be preferentially located on the main axes (horizontal, vertical, and diagonal), whereas the CoMs should be more evenly distributed in the random photographs. Using an arbitrary, but nevertheless strict criterion, a CoM was defined as being on an axis if it was within +0.5% of picture width/height from an axis (ie, including 1% of the total picture width or height). If CoMs were placed at random within the picture space, then about 4% of CoMs should meet that criterion. However, Figure 4 shows that CoMs are not distributed randomly, but rather the majority of CoMs are clustered around the middle of the image. Table 1 shows that 46.6% of 455 art photographs meet the criterion, compared with 39.7% of 595 random photographs,

a difference that is statistically significant (Chi-square = 5.018, 1df, $p = .028$, two-tailed). The mean distance to an axis for the art photographs was .0077 (SD .00789), compared with .0091 (SD .00870) for random images, a difference that is statistically significant using a t -test ($t = 2.557, 1050df, p = .011$), and a Mann–Whitney U -test ($p = .0041$).

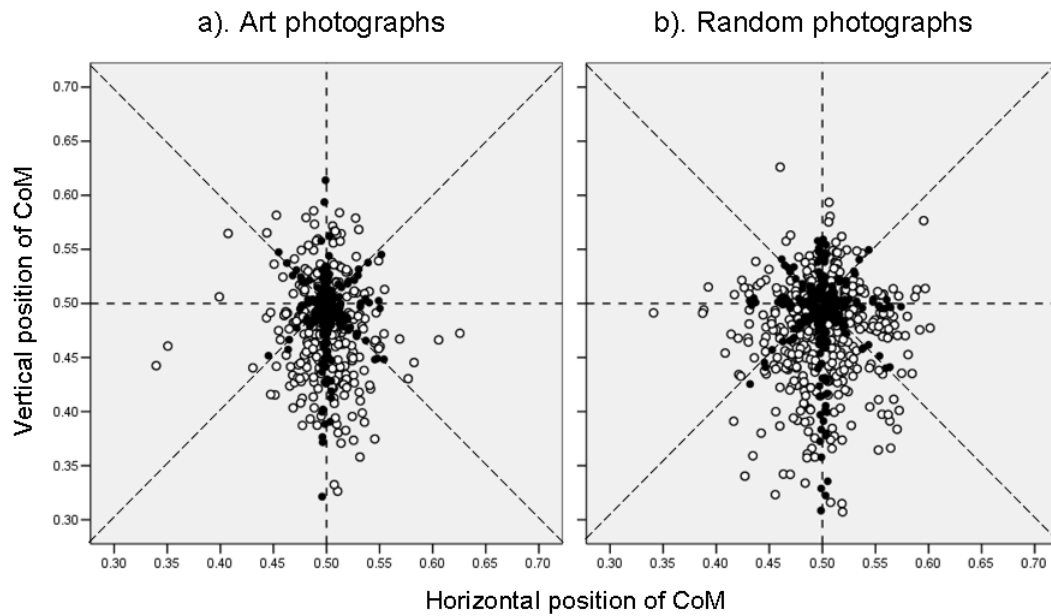


Figure 4. The CoMs of (a) the art photographs and (b) the random images. Points are plotted as solid black circles if they are within +0.5% of an axis or otherwise as open white circles. The horizontal, vertical, and diagonal axes are shown as dashed lines.

Table 1. Results of Study 1. 'On axis' indicates that the CoM was within 0.5% of a vertical, horizontal, or diagonal axis.

Nearest axis	The Photography Book		Random images	
	Original images	Internal controls	Original images	Internal controls
On axis	212 (46.6%)	57,186 (43.5%)	237 (39.7%)	75,400 (43.7%)
Off axis	243 (53.4%)	74,309 (56.5%)	360 (60.3%)	97,133 (56.3%)
<i>n</i>	455	131,495	597	172,533

2.3.2 *Statistical analysis of internal controls.* Table 1 shows that for the art photographs the CoMs of the sub-images used as internal controls are less likely to be located on an axis (43.5%) than is the case for the original images (46.6%). Statistical analysis of that result is not, however, straightforward, as the sub-images are not independent. The Arnheim–Ross model predicts that the CoM should be closer to an axis for the original images (but does not say which axis the CoM should be closer to, so that, for sub-images, the distance is taken to the nearest axis, whether or not that is the same axis as for the original image). For the j th of the i original images, let the distance to the nearest axis be d_j . For each of the k control images for the j th original image, let the nearest axis be at distance ${}_j p_k$. The variance of ${}_j p_k$ differs greatly between images and is very skewed, making analysis difficult. As a result, for

each original image, j, d_j is expressed as a percentile, q_j , of the distribution of values of $j p_k$ (so that on average the expected value is 50%, with a value of q_j less than 50% indicating that d_j is closer to an axis than the control values). As an example, it can be noted that for the image in [Figure 3](#) the CoM for the original image is on the 52nd percentile for the sub-image controls, meaning that it is marginally further from an axis than the set of sub-images. A one-tailed t -test can then be used to test whether the distributions of values of q_j across a set of images is significantly less than 50%.

2.3.3 Distance from CoM to nearest axis. The Arnheim–Ross model predicts that the CoM in art photographs should be closer to an axis than in internal controls derived from the image. The shortest perpendicular distance to any of the four axes was calculated and expressed, as above, as a rank of the shortest perpendicular distance for the internal controls. The average percentile rank for the art photographs was 46.32% (SE = 1.34), which is significantly different from 50% ($t = -2.733, 454df, p = .007$). In comparison, the average percentile rank for the random photographs was 52.31% (SE = 1.23), which is larger than 50%, although not significantly different from it ($t = +1.876, 596df, p = .061$).

2.3.4 Distance from CoM to image centre. A potential difference between the Arnheim and Ross theories is that Arnheim considers distance to the nearest axis whereas Ross considers distance to the image centre (so that balance would be the same whatever the orientation of the image). The distance of the CoM to the centre must of necessity be greater than to the nearest axis, and therefore we calculated distance to centre separately. In the original art photographs the CoM was a mean of .0444 (SD .0324) picture width/heights from the centre, compared with .0516 (SD .0365) picture widths in the original random photographs, a significant difference ($t = 3.326, 1050df, p = .00091$; Mann–Whitney U -test, $z = 3.248, p = .0012$). The distances to the centre in the original images and the internal controls were compared using the percentile method described earlier. The original art photographs were on the 46.52th percentile (SE 1.41; $t = -2.46, 454df, p = .015$), whereas the original random photographs were on the 54.42th percentile (SE 1.16, $t = +3.82, 596df, p = .00015$). Art photographs therefore have a CoM closer to the centre than do random photographs, both by direct comparison and in relation to internal controls.

2.3.5 Differences between horizontal, vertical, and diagonal axes. Although Arnheim implies that all four axes are important, that does not mean that CoMs are equally close to the four axes. [Table 2](#) shows the location of the nearest axis for the original images and the internal controls. The internal controls, both for art photographs and random photographs, show a very clear pattern, with a slight excess with the CoM closest to the vertical axis, probably reflecting the fact that many objects, particularly in a man-made world, have mirror symmetry around a vertical axis. Excluding cases where the CoM is closest to the vertical axis, there is little difference in the proportion of cases which are closest to the horizontal axis or either of the diagonal axes, suggesting that in general only the vertical axis is privileged. The original versions of the random photographs, perhaps not surprisingly, show a very similar pattern to the sub-images (and, of course, if the focal length had been different, they could have been random sub-images), with a predominance only of the vertical axis. In contrast, the original art photographs show a significantly different pattern from the random original photographs (Chi-squared = 29.58, 3 $df, p < .001$), and also the control sub-images, with a substantial excess of CoMs closest to the vertical axis, a deficit of images with a CoM closest to the horizontal axis, and with the proportions closest to a diagonal axis being similar to the random photographs and all of the control sub-images.

2.3.6 Historical shifts. A prediction from the Arnheim–Ross model might be that good composition, as expressed through balance, would have to be learned. Particularly for

Table 2. Results of Study 1. 'On axis' indicates that the CoM was within 0.5% of an axis.

Nearest axis	The Photography Book		Random images	
	Original images	Internal controls	Original images	Internal controls
Vertical	217 (47.7%)	42,126 (32.0%)	220 (36.9%)	55,185 (32.0%)
Horizontal	46 (10.1%)	27,823 (21.2%)	132 (22.1%)	40,793 (23.6%)
Diagonal 1	103 (22.6%)	32,710 (24.9%)	126 (21.1%)	37,873 (22.0%)
Diagonal 2	89 (19.6%)	28,836 (21.9%)	119 (19.9%)	38,682 (22.4%)
<i>n</i>	455	131,495	597	172,533

photography, where there was initially no experience of estimating balance in photographs, it might be expected that there would be a learning process over time, so that later images should be more balanced than earlier images (ie, a negative correlation with date). We tested that possibility by looking at the correlation between distance to the nearest axis and the date of production of a photograph. Based on 452 photographs, the correlation was $-.028(p = .557)$.

2.3.7 Non-linear modelling. The modelling so far has been based on the simple linear model of equation 1 in which the turning moment of the *i*th mass, m_i , at a distance, d_i , from the fulcrum, is simply $m_i \cdot d_i$, and hence averaged across all *i* masses, balance occurs when $\Sigma(m_i \times d_i) = 0$. Although the equation works in physics, it may not be appropriate for psychophysics, where the mass, *m*, and the effect of distance, *d*, may not scale linearly. A more general equation is therefore, $\Sigma [f(m_i) \times g(d_i)] = 0$, where mass is a non-linear function, $f(\cdot)$ of *m*, and distance is a non-linear function, $g(\cdot)$ of *d*. Clearly, there is an infinite number of such functions, and therefore here we will merely consider two intuitively reasonable possibilities, in order to assess their possible influence on CoM and their implications for the Arnheim–Ross model. It should also be noted that, unlike the equation $\Sigma (m_i \times d_i) = 0$, the equation $\Sigma [f(m_i) \times g(d_i)] = 0$ cannot readily be converted to a form such as that in equation 1, and in the general case it has to be solved iteratively rather than analytically.

For photographs the obvious non-linear transformation of *m* is the gamma correction function, which is much used in digital image analyses, where $m' = m^\gamma$, and intensities are scaled in the range 0 to 1. When $\gamma = 1$, the image is unchanged; when $\gamma > 1$, the image is systematically (but non-linearly) darkened; and when $\gamma < 1$, the image is lightened overall. For the effect of distance from the fulcrum, an equivalent non-linear transformation is $d' = d^p$, where the effect of *m* is proportional to the *p*th power of its distance from the fulcrum (and we refer to *p* as the 'distance power'). For $p > 1$, distant masses have a proportionately greater effect on the turning moment than when $p = 1$. A straightforward model would therefore seem to be:

$$\sum_{i=1}^n m_i^\gamma \cdot r_i^p \tag{2}$$

A moment's reflection, however, reveals that there is a problem with this model, as the values of d_i are usually distributed to the right and the left of the fulcrum, with some having negative signs and others positive. However, if these values are squared (ie, $p = 2$), then all

are positive, and the equation cannot be solved. A solution is to raise the absolute value of d_i to the power p , and multiply it by the sign of d_i , as shown in equation (3).

$$\sum_{i=1}^n (m_i^\gamma \cdot \text{sign}(d_i) \cdot |d_i|^p) = 0 \quad (3)$$

In order to keep the problem tractable, the model has been fitted to the images from *The Photograph Book* for parameters $\gamma = 0.5, 1$ and 2 , and $p = 0.5, 1$ and 2 (where $\gamma = 1$ and $p = 1$ is equivalent to equation 1). Without giving examples, suffice it to say that the effect of such gamma corrections on the photographs is visually quite dramatic (and for examples of gamma corrections in general, see http://en.wikipedia.org/wiki/Gamma_correction). Gamma corrections to the images were applied to entire images using the `imadjust` function in Matlab, and equation 3 was then solved using the `fsolve` function.

Table 3 compares the effect of the three levels of γ and p for the 455 images from *The Photograph Book*. Columns 3 and 4 show the mean difference in position of the CoM compared with CoM for $\gamma = 1$ and $p = 1$. Altering γ and p hardly alters the location of the CoM, with means very close to zero (and typically less than .001—ie, one thousandth of picture width or height). Likewise the CoM correlates very highly with the X and Y positions of the CoM for $\gamma = 1$ and $p = 1$, all correlations being higher than .965. Finally, columns 6 and 7 show the mean percentile for the original image (after γ and p have been altered), compared with the internal control images. All percentiles are less than 48.66%, and in five of the nine cases, reach significance with $p < .05$. Overall, the conclusion has to be that the results are robust against quite large changes in γ and p .

Table 3. Analyses of effects of altering gamma and distance power on location of CoM for the 455 images from *The Photograph Book*. Note that when $\gamma = 1$ and p (Distance Power) = 1, then the analysis is equivalent to that presented previously.

γ	p	Mean (SD) distance of CoM from CoM for $\gamma = 1$ and $p = 1$		Correlation with CoM position for $\gamma = 1$ and $p = 1$	Percentile of original full image versus internal controls	
		Mean	SD		Mean percentile	One-tailed t -test versus 50%
.5	0.5	X: -.00016	X: .01384	X: .979	48.66%	$p = .318$
		Y: -.00792	Y: .02050	Y: .972		
.5	1	X: -.00015	X: .00745	X: .990	47.76%	$p = .097$
		Y: -.00477	Y: .01076	Y: .994		
.5	2	X: -.00011	X: .00468	X: .983	47.06%	$p = .026$
		Y: .00079	Y: .00623	Y: .991		
1	.5	X: -.00002	X: .00611	X: .993	47.83%	$p = .105$
		Y: -.00280	Y: .00951	Y: .989		
1	1	X: 0	X: 0	X: 1	46.32%	$p = .007$
		Y: 0	Y: 0	Y: 1		
1	2	X: .00005	X: .00712	X: .989	45.06%	$p = .0002$
		Y: .00451	Y: .01115	Y: .991		
2	.5	X: .00001	X: .00603	X: .979	47.44%	$p = .055$
		Y: .00364	Y: .00955	Y: .986		
2	1	X: .00006	X: .00847	X: .980	46.02%	$p = .004$
		Y: .00585	Y: .01374	Y: .985		
2	2	X: .00012	X: .01291	X: .965	45.32%	$p = .0009$
		Y: .00905	Y: .02094	Y: .969		

2.4 Study 1: Discussion

This descriptive study of a representative selection of high quality art photographs, which can be regarded as canonical, has a number of interesting findings. The art photographs were compared with external controls (a large set of random photographs taken by one of the authors) and a set of internal controls produced by sampling from within the image itself. In each case it was clear that art photographs have a CoM which is closer to an axis, and also closer to the image centre, than is the case for random images. All images show a tendency to be closer to the vertical axis than other axes (probably because many objects in the visual world show mirror symmetry about the vertical axis), but while random images show an equal distribution about the horizontal axis and the two diagonal axes, art photographs are less likely to have a CoM close to the horizontal axis.

A possible problem for Study 1 is that the reproductions of the images from *The Photography Book* are not exact. The original printed photographs have themselves been photographed for the books, they have been printed, and they have then been scanned. Each process has its own characteristics, described most clearly by its gamma function. A concern has to be that if the gamma functions change, then the CoM will itself also change. We examined this possibility before commencing the study and found that altering gamma across an entire image, within a reasonable range, produced hardly any change in the CoM at all. We therefore felt confident that our analyses conform to any reasonable differences in reproduction of the images. The analyses of Table 3 confirm that intuition systematically, large alterations in γ , and indeed alterations also of p , not substantially altering either the location of the CoM or the statistical significance of the results. Of course, in the extreme, alterations of γ and p must result in non-significant results. When $\gamma = 0$ or $\gamma = \infty$, then the image will be entirely black or entirely white, and the CoM must then be precisely at the centre both for the original image and the internal controls. Likewise, as p tends to 0, then the fulcrum is at the median of the masses, and as p tends to infinity, so all pixels to right and left of the physical centre equally influence balance, and the CoM must be at the centre. Finally, it should be emphasised that the alterations of gamma were applied uniformly across the entire image (in contrast to the 'gamma-ramp' experiment in Study 2).

Although not a strong test of the Arnheim–Ross model of balance, the present results are compatible with it, the CoM of recognised works of art being closer to a major axis than is the case in control images. Other processes could, however, result in that finding, particularly if, say, a large dark mass symmetrically fills most of the central area of a photograph (as in some portraits or still lifes), as then the CoM will be constrained to the image centre, and any sub-image will be unlikely to show such symmetry, and hence the CoM will be less close to the centre (and hence less close to an axis). Likewise, the random control photographs are less likely to contain such symmetrically placed central images.

A stricter test of the Arnheim–Ross hypothesis is therefore required, which uses experimental manipulation of the canonical images, so that the CoM is either on, or is not on, an axis.

3 Study 2: The Gamma–Ramp experiment

3.1 Study 2: Introduction

The Arnheim–Ross theory says that an image should be preferred more if it has its CoM on one of the major axes and should be preferred less if the CoM is off a major axis. Study 2 uses a simple method to move the CoM while retaining the content of the image.

The concept of gamma adjustment is much used in image processing, mainly for adjusting the properties of cameras, scanners, printers, and monitors, to take into account the differing non-linearities of the various devices, which otherwise make images appear too dark or

alternatively appear as washed-out. Photographers of the old school who have worked with paper and chemicals in darkrooms will also remember that printing paper used to come in various degrees of 'hardness', hard paper producing very contrasty, black-and-white prints, whereas soft paper has less contrast, so that the darker tones are more similar to the lighter tones. Gamma for computer-generated images is straightforward to adjust, and normally, as with hard or soft paper in a darkroom, a single value of gamma is chosen for the entire image. However, with a computer it is also possible to adjust gamma so that it varies linearly across an image, be it up-down or from left-right (or both). If the gradient is kept small, the effects are subtle, but a consequence is that the centre of mass moves on or off of axes, according to the values that are chosen. In effect, the image is being printed on paper which is harder or softer in some parts than others.

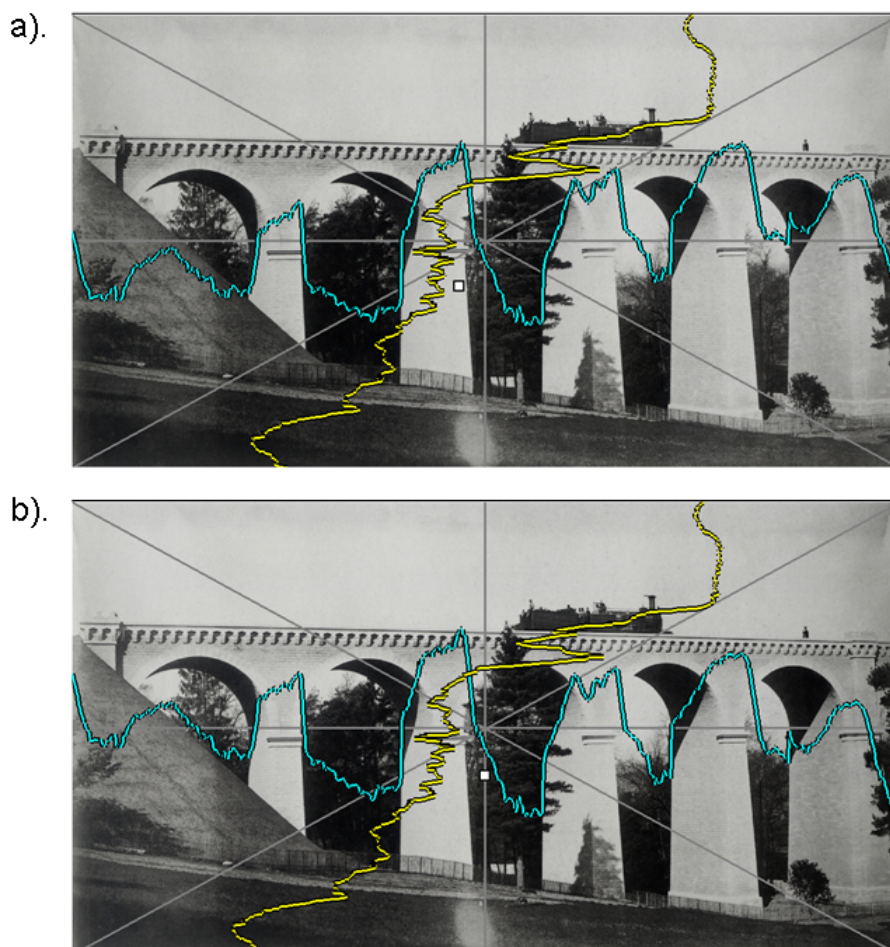


Figure 5. The bridge at Chantilly by Edouard-Denis Baldus (1855). Overlaying the image are the horizontal, vertical and diagonal axes (in grey), and the centre of mass (shown as a white square). The blue line shows the lightness of the columns (light up and dark down), and the yellow line the lightness of the rows (light to right and dark to left). (a) Original photograph in which the centre of mass is between axes. (b) Gamma ramps have been imposed vertically and horizontally, which have moved the centre of mass so that it is the same standardised distance from the image centre but is now on the vertical axis. Most of the adjustment has been horizontally (blue line), and a careful comparison of the two images will show subtle differences, which are responsible for moving the centre of mass. Note: This image was not one of those actually used in Study 2 but is used here as an illustration since it is in *The Photography Book*, though its date is out of copyright (see http://en.wikipedia.org/wiki/File:%C3%89douard_Baldus_-_Paysage_Pris_du_Viaduc_de_Chantilly.jpg).

The effects of the gamma-ramp adjustment can be seen in [Figure 5](#), where at the top is the original version of an art photograph in which the CoM is between axes, and at the bottom is a modified version in which the CoM has been moved so that although remaining at the same standardised distance from the centre, it is now on the vertical axis. Comparison of the horizontal weights (the blue lines) will show that the left side in [Figure 5b](#) is slightly lighter and the right side is slightly darker than in [Figure 5a](#), which has shifted the CoM to the right and on to the vertical axis. There are also small changes vertically to ensure the CoM is still the same distance from the image centre.

In Study 2 the CoM of art photographs has been moved slightly, typically by about 20 or so degrees, keeping the distance of the CoM from the image centre constant, but either moving the CoM from an axis so that it is instead midway between axes, or alternatively moving a CoM which is midway between axes in the original so that it is precisely on an axis. The Arnheim–Ross model should predict that preference will be higher for images which have the CoM on an axis rather than off it.

Examination of the two images in [Figure 5](#) shows that they are extremely similar, to the point that, in a preliminary study, it was found that a conventional paired-comparison design, with the images placed side-by-side, was extremely hard for participants to carry out, the two images appearing almost indistinguishable. However, it was also found that if the two images were overlaid and the viewer swapped back and forth between the two images, then the differences became much clearer. A successive paired comparison design was therefore developed in which one of the images randomly came upon the screen, and then at each press of a key the two images swapped. Participants were allowed to alternate as many times as they wished before they indicated which of the pair they preferred.

3.2 Study 2: Method

3.2.1 Stimuli. Twenty images were chosen from *The Photography Book* in which the CoM of the original image was on an axis ('On') and was a moderate distance away from the image centre. Gamma ramps were then used to produce a modified image ('OnToOff'), in which the CoM was moved so that it was off-axis, being midway between the original axis and the next axis around (half clockwise and half anti-clockwise), keeping constant the distance of the CoM from the image centre. Twenty original images were also used in which the CoM of the original image was midway between axes ('Off'), and was also a moderate distance from the image centre, and gamma ramps were used to modify those images so that the CoM was on an adjacent axis ('OffToOn'), randomly clockwise or anti-clockwise. The CoM was moved by a special purpose Matlab program, which used an optimisation procedure to find the appropriate gamma ramps in order to move the CoM to the required position.

3.2.2 Procedure. Participants were asked to compare the original version of an image with its modified version (ie, On with OnToOff or Off with OffToOn). A successive paired comparison design was used (see above), and when participants had decided which of the two images they preferred, they then indicated by how much they preferred it to the other image on a three-point scale (Very much, Quite a lot, Not much).

Participants. A total of 38 participants took part, who had also taken part in Study 6 of McManus et al (2011). Ten participants were experts (8 male, 2 female; mean (median) age = 26.8 (28.5), SD=3.12), all of whom had attended art school and were professionally involved in the visual arts. Twenty-eight participants were non-experts and were a convenience sample, mostly but not entirely consisting of undergraduates at University College London (4 male, 24 female; mean (median) age = 25.89 (21), SD= 11.11). No subjects were paid for their participation.

Ethics. Ethical approval for this and the other studies was given by the Ethics Committee of the Department of Psychology of University College London.

3.3 Study 2: Results

Although participants in effect made judgements on a six-point scale, for simplicity of analysis the responses were treated as binary, as a preference for one of the stimuli in a pair. The sequential paired comparison method worked well, with participants making an average of 4.8 alternations on each trial, looking at the stimuli for an average of 7.3 s before making a preference.

3.3.1 Preference for On-Axis images. Each participant made 40 judgements, with one member of each stimulus pair having the CoM on-axis and the other having it off-axis. Table 4 shows that of a total of 1,520 judgements, 736 (48.4%) were for the on-axis image and 784 (51.6%) for the off-axis image. The mean percentage of on-axis preferences by the 38 subjects was 48.42% (SE = 1.14%), which is not significantly different from a chance expectation of 20 ($t = -1.384, 37df, p = .175$). Neither was there any significant difference in the percentage of times the on-axis image was preferred by the experts (mean = 47.75%, SE = 2.06%) or the non-experts (mean = 48.66%, SE = 1.28%; $t = .347, 36df, p = .730$). Of the 38 participants, only 2 (5.3%) reached the $p < .05$ level of significance individually (means = 32.5% and 70.0% correct), which is what would be expected by chance with 38 participants.

Table 4. Results of Study 2, expressed in two ways, (top) in relation to whether the on axis or off axis image was preferred, and (bottom) whether the original or the modified image was preferred.

	Preferred Off Axis	Preferred On Axis
Original Off Axis	507 (66.7%) 'Off'	253 (33.3%) 'OffToOn'
Original On Axis	277 (36.4%) 'OnToOff'	483 (63.6%) 'On'
Total	784 (51.6%)	736 (48.4%)
	Preferred Modified Image	Preferred Original Image
Original Off Axis	253 (33.3%) 'OffToOn'	507 (66.7%) 'Off'
Original On Axis	277 (36.4%) 'OnToOff'	483 (63.6%) 'On'
Total	530 (34.9%)	990 (65.1%)

3.3.2 Preference for original images. In each pair, one image was an original and one was modified. Table 4 shows that overall 990 out of 1,520 preferences (65.1%) were for the original image, and 530 (34.9%) for the modified image. Overall subjects made a mean of 26.05 preferences for original images (SE = .62), which is significantly different from a chance expectation of 20 ($t = 9.84, 37df, p < .001$). Experts made 28.70 (SE .62) preferences for original images, compared with 25.11 (SE .73) preferences by non-experts ($t = 2.80, 36df, p = .008$). Nineteen participants were significantly better individually than chance with $p < .05$, whereas none were significantly worse than chance (range 67.5% to 80.0%); the group significantly better than chance included 9 of the 10 experts.

3.3.3 Analysis by images. Analysis at the level of the 40 individual photographs showed that overall there was no overall preference for the images on-axis (mean percentage = 48.42%; SD 23.25%; $t = .429, p = .670$). However, considering each photograph individually, 10 (25.0%) showed preferences that were significant ($p < .05$) for the on-axis image, and 11 (27.5%) were significant for the off-axis image. These differences in fact reflect differences in preference for the original image. Overall there was a highly significant preference for original images (65.13% correct, SD = 17.57%, $t = 5.448, 39df, p < .001$), with preferences for the original that were individually significant with $p < .05$ for 20 (50.0%) of the images (65.8% to 97.4% correct). For one photograph only (2.5%) was there an individual significance at the $p < .05$ level of significance for the non-original image, and that can probably be put down as a chance phenomenon.

3.4 Study 2: Discussion

The manipulation of gamma ramps provides a strong, experimental test of the Arnheim–Ross model, in large part because the images are very similar—indeed, differing almost entirely in the location of the CoM (to the extent that a simple paired comparison is probably too difficult for most participants). The sequential paired comparison design worked well, but there was no evidence that participants, be they expert or non-expert, had any preference for images where the CoM was on-axis rather than off-axis. However, and it is an important methodological and theoretical point, there were clear preferences for original over modified images, and that was despite the differences between images being quite subtle, and, of particular interest, the preference was stronger in the experts compared with the non-experts. The implication is that, even if it is not the CoM that is the crucial aspect of the organisation of an image, there has to be some property that is being perceived by the participants and which is telling them that an original image is preferable to a modified image.

4 Study 3: Cropping of photographs

4.1 Study 3: Introduction

An important feature of Study 1 was that it used photographs which are of recognised aesthetic quality. A major problem, however, is that in most cases one has little sense of what the image might have looked like had the photographer chosen to take a slightly different photograph, moving the camera up or down, to right or left, and so on. There are, of course, occasional exceptions, as when one can examine the contact prints of photographers such as Diane Arbus (Lee 2003) or Henri Cartier-Bresson (2006), but mostly we have no sense of the choices made by a photographer. Study 3 used a different method, based on the cropping paradigm which has been described elsewhere (McManus et al 2011), in which participants see a large image and then crop it down to half of its size (a quarter of its area). The psychometrics of the method have been described in detail in the previous paper. Study 3, which re-analyses data collected previously, asks whether the cropped images that are chosen have a CoM that lies closer to a major axis than would be expected by random cropping. The Arnheim–Ross model predicts that they should.

Figure 6 shows an example of one of the images used in the cropping experiment. The small inner yellow box is the Inclusion Box, which must be included in any cropping. As a result, the crop window cannot be outside the area delineated by the outer yellow box, and the midpoint of any possible crop window cannot be outside the middle yellow box. The individual red and pink squares are the centres of the crop windows for non-experts and experts, respectively.

4.1.1 *The balance loci.* The green, blue, purple, and pink lines are complicated, but each indicates a 'balance locus' at which for a sub-image centred at that point, the image would be balanced around the vertical axis (green), horizontal axis (blue), major diagonal (purple), or minor diagonal (pink). Thus, for instance, consider the green line. If a sub-image, of size 512 x 384 pixels, had its centre anywhere on this green line, then the sub-image would be balanced precisely around the vertical axis (but not necessarily around the other axes). The orange box indicates the area of all possible image midpoints, irrespective of whether the Inclusion Box is included, and the balance loci are plotted for that entire area. Where the four lines for the separate balance loci intersect is the location of the centre of a sub-image which is balanced around all four axes, and that sub-image necessarily has its CoM precisely at its physical centre. Note that there can be one, several, or no points in an image at which all four balance loci intersect, and in Figure 6 there are two such locations, although the one towards the upper-left corner would not be legitimate for the cropping experiment, given the

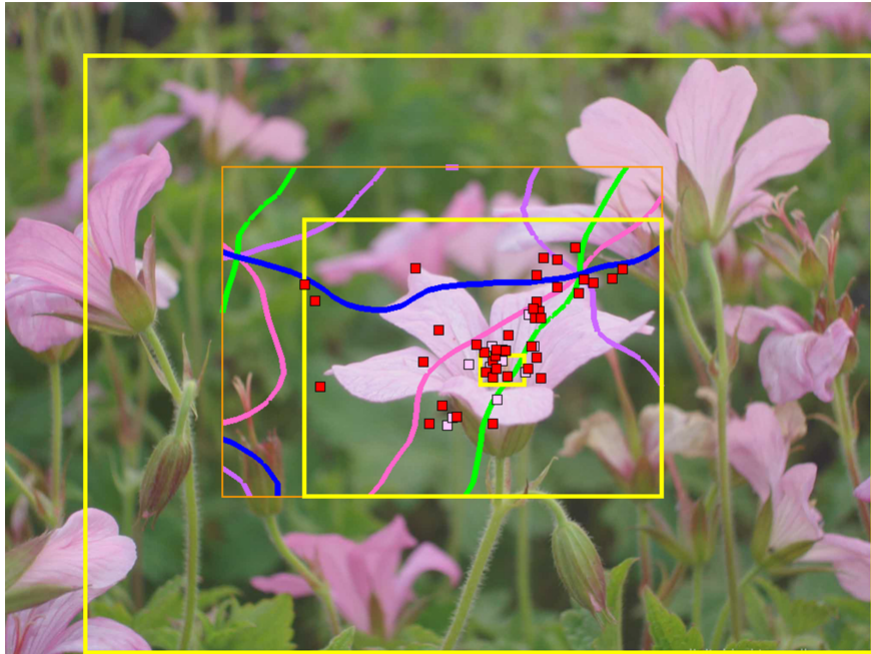


Figure 6. An example of the results of the cropping experiment from Study 3. The small yellow box is the Inclusion Box, the outer yellow box is the limits for a crop, given the size of the inclusion box, and the middle yellow box is the limits of the midpoints of the crops, given the Inclusion Box. The red and pink squares are the centres of the crops made by the individual participants (red: non-experts; pink: experts). The blue, green, purple, and pink lines indicate the loci of the centres of crops, which would have their CoM on the vertical, horizontal, major diagonal, and minor diagonal axes, respectively. The orange box shows the limits for the midpoints of all possible images, and the balance loci are also shown for all possible midpoints, including those not permitted by the Inclusion Box.

location of the Inclusion Box. Later, in [Figure 11](#), will be seen images in which there is only a single balance locus, and hence there is no possible intersection for all four balance loci.

4.2 Study 3: Method

The data re-analysed here are described more fully in Study 4 of McManus et al (2011), and only a brief description is necessary here.

4.2.1 Procedure. In the viewing phase, participants firstly saw a photograph of size 1024 x 768 pixels, which filled the screen of a computer monitor. They were allowed to look at the photograph for as long as they liked, and then a click of the mouse button moved to the cropping phase in which a region of the photograph of size 512 x 384 pixels was visible to the participant. Moving the computer mouse moved the visible region, and the participant had as long as they wished in order to explore possible locations for the crop, clicking the mouse when the image 'looked best'. In each image there was an 'Inclusion Box' – a region of the image which can be regarded as the 'subject' of the picture, and which the software ensured was included in any crop. As a result not all positions of the cropping window on the screen were possible.

4.2.2 Stimuli. Although participants saw 80 images in the original study, some images were repeated, and others were presented in variant forms such as in monochrome or as thresholded image. Only the first 30 images will therefore be analysed here, which were all in full colour and being seen for the first time.

4.2.3 *Participants.* Of 51 participants, 10 were experts in the visual arts, having degrees in the visual arts and being on the MA course in photography at the Royal College of Art. The other 41 participants were non-experts, typically undergraduates at University College London.

4.3 Study 3: Results

A similar analysis was carried out to that used in analysing art photographs in Study 1. For each cropped image produced by each subject, the distance of the CoM to the nearest axis was calculated. A set of control images was generated, which systematically sampled the range of possible crop window positions, taking into account the constraints of the Inclusion Box, and for each the distance of the CoM to the nearest axis was calculated. The distance to the nearest axis for the actual crop window was calculated as a percentile of the distribution of distances for all possible crop windows, with lower percentiles meaning that the actual crop window was closer to an axis than was the case in the control images.

Considering all 1,530 croppings, the CoM was on average 5.86 pixels from an axis (SD 5.40), compared with an average of 10.70 pixels (SD 8.22), for all of the control images. Those values cannot be directly compared statistically, as they are not independent of one another. On average the 1,530 cropped images were on the 30.25 percentile (SD 24.98, SE 0.69) compared with the control croppings. The percentiles for each participant were averaged across all of the 30 croppings, giving a mean percentile of 30.25 (SD 4.13, SE .58), which is significantly different from 50% ($t = -34.18, 50df, p < .001$). The difference between experts (mean = 29.17, SE 1.29) and non-experts (mean = 30.56, SE = .65) was not significant ($t = .986, 49df, p = .329$). The range of subject means was surprisingly narrow, from 20.29 to 39.21, with all being significantly different from 50% with $p < .05$.

An analysis was also carried out at the level of the 30 images, the percentile being calculated across all 51 subjects. As before, the mean was 30.25, but with SD = 14.89 and SE = 2.72, the mean being statistically different from 50% ($t = 7.26, 29df, p < .001$). However, examination of the individual means for each image found that although 47 of the 51 were significantly less than 50%, 2 were not significantly different from 50% (49.65% and 51.76%), and two were significantly greater than 50% (65.20% and 70.76%). The image shown in [Figure 6](#) has a mean percentile of 18.56.

4.4 Study 3: Discussion

Study 3 shows that participants in a cropping study, in which a fixed image is cropped to half of its size, produce crops whose CoMs are closer to an axis than would be the case for random croppings of the same image, and that is compatible with the Arnheim–Ross model. The result is somewhat stronger than the similar analysis with the art photographs, not only because the effect is much larger in percentile terms, but also because all of the control images were possible choices for the participants in the experiment. However, as with Study 1, the result provides only weak support for the Arnheim–Ross model itself, since it is possible that other factors (such as preferences for overall symmetry, etc) are producing the result, rather than it being dependent on the relationship of the CoM to an axis. That possibility is seen in [Figure 6](#), where a number of the crops are centred on the centre of the flower, and hence are close to several of the axes. It is also of interest that experts do not produce crops that are closer to an axis than do non-experts.

5 Study 4: The Spiders-Web experiment

5.1 Study 4: Introduction

Although the data on cropping, in Study 3, are compatible with the Arnheim–Ross model, they are not a strong support for it, as many other factors may differ between the cropped images and the control images. Study 4, which we call the Spiders-Web experiment, is, like

Study 2, the gamma-ramp experiment, designed to control many of the differences that may occur between a cropped image and the controls used in Study 3 and to produce paired images which differ almost entirely because one has a CoM on an axis and the other does not.

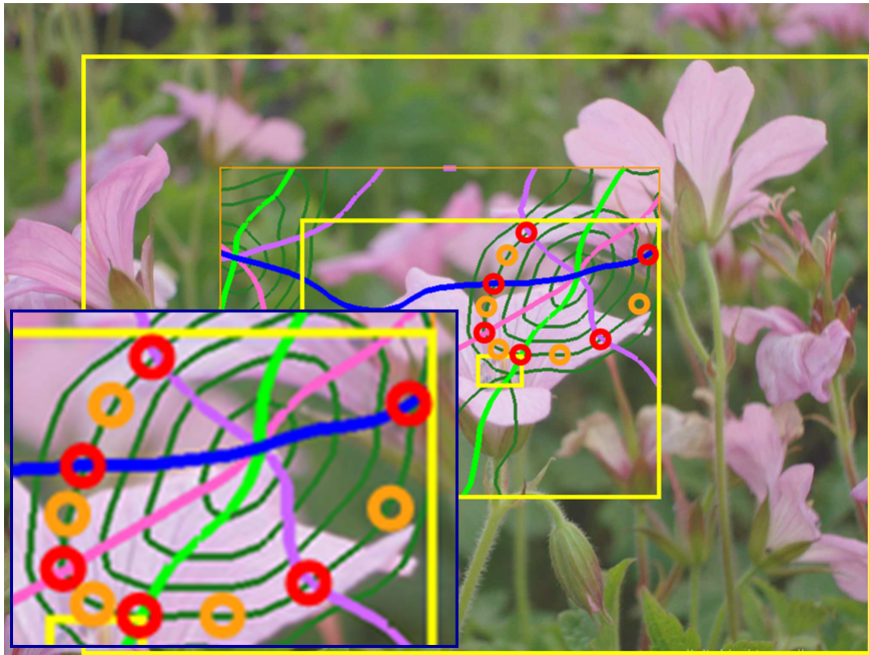


Figure 7. The design of the Spider-Web experiment. See Introduction to Study 4 for details. The image in the bottom left-hand corner shows an enlargement of the key area.

The design of the experiment can best be understood by looking at [Figure 7](#). As in [Figure 6](#), the image shows the green, blue, purple, and pink lines which are the balance loci for the centre of a cropped image which has its CoM on one of those four axes. The dark green lines in [Figure 7](#), which are similar in some ways to a spider's web, and hence the name, are the loci of balance points on individual balance loci which are equidistant from the physical centre of the sub-image. Consider just one of those dark green lines, for which the intersections with the green, blue, purple, and pink lines have been indicated by red circles. The images corresponding to these red circles all have a CoM on precisely one axis, and those CoMs are equidistant from the physical centre of the corresponding sub-image. In contrast, the orange circles indicate the centres of sub-images which are the same distance from the physical centre of the corresponding sub-image as are the red circles, but are not on any axis, and are indeed exactly midway between any two axes. The prediction for the Arnheim–Ross model is that there should be a preference for images based on the red circles compared with those based on the orange circles. If adjacent red and orange circles are used in a pair, then the content of the images should differ only very slightly, meaning that the preference should be a preference based almost entirely on composition, and particularly on balance, rather than on content.

5.2 Study 4: Method

5.2.1 Procedure. A conventional paired comparison experiment was carried out in which, as a part of a larger preference experiment, participants saw a series of 16 pairs of images which used the Spider-Web method, to make one image in each pair have its CoM on an

axis, while the CoM of the other image was between adjacent axes. Participants responded on a 6-point scale (Strong, Medium, or Weak preference for the left-hand image or Weak, Medium, or Strong preference for the right-hand image). Participants were allowed to work at their own pace.

5.2.2 Stimuli. Examination of the loci for balance on the four axes in a large number of images used in previous experiments found that many images do not show loci for balance on all four axes, and as a result have no point at which the exact mid-centre of the image is also at the CoM. Four images were identified from the set used in previous experiments in which there was a four-way intersection, and it was also possible to have a spider's web in which the contour of equidistant from the centre was sufficiently far from the intersection. Four pairs of images, one on-axis and the other off-axis, were then generated from each of these four images, making 16 pairs altogether. On-axis members of the pair were presented randomly to right or left in each pair.

5.2.3 Participants. The participants in this study were the same 38 participants (10 experts, 28 non-experts) as in Study 2.

5.3 Study 4: Results

The 6-point preference judgements, for ease of analysis, were reduced to 2-point preferences, for either the on-axis or off-axis image. Overall, 608 preferences were made, of which 259 (42.6%) were for on-axis images in each pair, and 349 (57.4%) were for off-axis images. Each of the 38 participants made 16 preferences, with a mean of 42.6% for the on-axis image (SD 12.4%; SE 2.01%), which is significantly less than the 50% expected by chance ($t = 3.679, 37df, p = .00074$). Overall, the 10 experts preferred the on-axis image on 39.4% of occasions (SD 13.2%), compared with the non-experts on 43.8% of occasions (SD 9.8%), a non-significant difference ($t = .956, 36df, p = .345$).

5.4 Study 4: Discussion

The Spider-Web experiment provides no support for the Arnheim–Ross theory, as in images which are minimally different apart from one image having the CoM on-axis and the other off-axis, there was no preference for the on-axis image. Indeed, it is striking that there is a significant preference for the off-axis image. The reasons for that are far from clear, but while not supporting the Arnheim–Ross theory, they do suggest that the participants are making meaningful judgements of some sort, and that there is agreement between the participants. What the features are which produce that preference are not obvious at present.

In many ways the result of Study 4 is similar to that of Study 2. In each case there was a weak result (in Study 1 and Study 3) which supported the Arnheim–Ross model (but could have been due to other factors as well), but then a much more focussed experiment, in one case using the Gamma-Ramp, in the other the Spider-Web, found no support for a very specific prediction from the theory.

Taken together, Studies 2 and 4 suggest that at least when looking at complex photographic images, there is no preference for images which are balanced in the sense of the physicalist model proposed by Ross and Arnheim. At that point one has to wonder whether the model even works for much simpler images of the type that Arnheim uses in chapter 2 of *Art and Visual Perception* of two black disks placed in a square frame, and that is investigated in Study 5.

6 Study 5: Cropping of Arnheim's own geometric stimulus

6.1 Study 5: Introduction

Arnheim, in Figures 5a and 5b of *Art and Visual Perception* (p. 18) (reproduced here as [Figures 8a and 8b](#)), shows two square frames, the first of which (a) has two black disks arranged along

the main diagonal, symmetrically around the mid-point. In the second frame (b) the same two disks are moved so that one is now at the exact centre of the image and the other further along the diagonal, about midway between the midpoint and the top right-hand corner. After comparing them, Arnheim goes on to say that:

Like a physical body, every finite visual pattern has a fulcrum or center of gravity. And just as the physical fulcrum of even the most irregularly shaped flat object can be determined by locating the point at which it will balance on the tip of a finger, so the center of a visual pattern can be determined by trial and error. According to Denman W. Ross, the simplest way to do this is to move a frame around the pattern until the frame and pattern balance; then the center of the frame coincides with the weight center of the pattern (p. 19).

Study 5 does exactly what Arnheim suggests, both with his figure (and the two versions are equivalent once the frame can move around the two disks) and with a range of related images. Arnheim considers only a square frame, but we also consider rectangular frames, and in addition make the two disks of different sizes, make one of them grey, and also place a light-grey/dark-grey ramp horizontally or vertically behind the disks. All of these manipulations would be expected to alter the location of the CoM, and hence the optimal position of the frame. We also investigate right-left and top-down asymmetries since Arnheim suggests that each of them will alter the apparent balance of an image (Arnheim 1974, pp. 30 & 33). A point not discussed by Arnheim, despite most painted and photographic images having rectangular rather than square frames (Fechner 1876), is that what Palmer and Guidi (2011) call the 'structural skeleton' of a rectangle is more complex than that of a square. As Palmer (1991) suggests, for a rectangle, there is global symmetry around the horizontal and vertical midlines, but there is also an area of 'local symmetry' which runs at 45 degrees from the corners and does not intersect with the centre of the rectangle.

6.2 Study 5: Method

The method follows the cropping procedure described elsewhere (McManus et al 2011), and also briefly described above for Study 3.

6.2.1 Procedure. The procedure was similar to that for Study 3. Participants firstly, to ensure that they understood the nature of the cropping task using real photographs, cropped a series of 40 photographic images, including ten repeats, that have been used in several previous studies. In these cases the overall image size was 1024 x 768 pixels, which filled the screen of a computer monitor, and the crop window was 512 x 384 pixels. Subjects then saw a series of 24 abstract images similar to that described by Arnheim and were asked to crop these in a similar way to that in which the photographs were cropped. The first 20 images also used a crop window of 512 x 384 pixels, and the last four used a square crop window, of size 384 x 384 pixels, in order to reproduce Arnheim's Figure 5a.

6.2.2 Stimuli. The stimuli were divided into four sets, the members of which were presented in random order, with the exception of the four images with a square crop window, which were presented last. The four sets are shown in Figure 8 to Figure 11. In each case the Inclusion Box was set so that the two large disks set at 45 degrees had to just be included. The figures also show the location of the balance loci along the vertical, horizontal, and diagonal axes. The sets consisted of:

1. Square images similar to those used by Arnheim (Figures 8c–f). Figure 8c replicated the Arnheim stimulus as closely as possible. For a 384-by-384-pixel square frame the disks were of diameter 149 pixels and were 205 pixels apart. The black disks had a grey level of 0. As in Arnheim, the disks are on a 45-degree diagonal. A point of some interest in Arnheim is that he considers a square frame so that the diagonal is at 45 degrees. However, since most of our work has used a frame with an aspect ratio of 4:3,

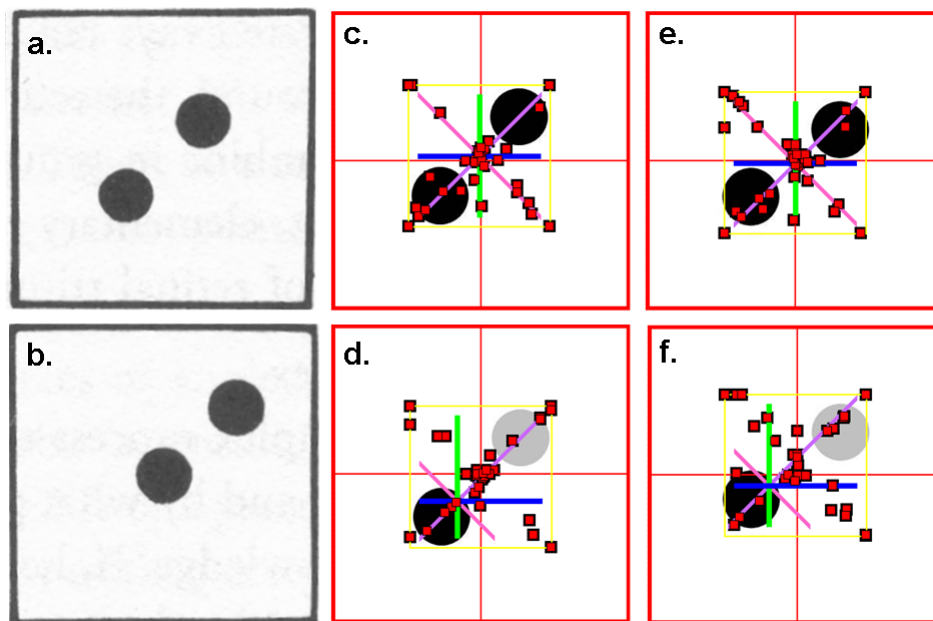


Figure 8. Panel (a) and (b) show the simple images described by Arnheim, with a square frame and two black disks. Panels (c) to (f) show the results of Study 5, in which the participants followed the procedure described by Ross, and placed the frame around the disks until the position looked best. Note that unlike Figure 6, in which one sees the centres of all frames in relation to the entire set of possible positions on the screen, Figures 8, 9, 10, and 11 reverse the process and show just the average cropping window (see text) in which the preferred placings of the disks are shown along with the individual preferences relative to that average. What one sees is therefore what a typical participant would say is the best location of the disks within a frame. As in Figure 6, the blue, green, purple, and pink lines indicate the loci of the centres of crops which would have their CoM on the vertical, horizontal, major diagonal, and minor diagonal axes, respectively. If on average the participants have a preferred position that is on an axis, then the intersection of the red cross-hairs should be on a coloured line.

for which the diagonal is at an angle of 36.9 degrees, Figure 8e uses the same two disks with that angle. Note that having the disks at an angle of 36.9 degrees does not alter the predictions of the model (shown as the balance loci) because the frame is square. Figures 8d and 8f are similar to Figures 8c and 8e, except that the upper disk has a grey level of 192 (black = 0; white = 255). A prediction of the Arnheim–Ross theory is that having one light grey disk should shift the balance point (and the intersection of the balance loci in panels d and f have shifted towards the black disk, and is also now no longer the same in Figures 8d and 8f).

2. Anisotropy of left-right and up-down. Arnheim suggests that objects in images are heavier at the top of the image and on the right of the image. Figure 9 shows four images in which there is a standard black disk, and also a grey disk (grey level 192) which is placed at top left (Figure 9a), top right (Figure 9c), bottom left (Figure 9b), or bottom right (Figure 9d). If anisotropies exist, then right and left and top and down should not be mirror images of one another (although they are isotropic in the model).
3. Variation in size and greyness of one of the disks. A smaller disk should weigh less than one that is larger, as also should one that is visually lighter. Figure 10 shows 12 stimuli which systematically varied the size and greyness of the upper left-hand disk. The disks on the top row (a, d, g, j) are of standard size; those on the middle row (b, e, h, k) have half the area of a standard disk (and hence the diameter is $1/\sqrt{2}$ of the standard disk). The disks on the bottom row (c, f, i, l) have a diameter half that of the standard

disk (and hence a quarter the area). The disks in the first column are all black (grey level = 0), whereas the upper disk in columns two, three, and four have grey levels of 64, 128, and 192. As the upper disk gets smaller and lighter-grey, so the intersection of the balance loci moves towards the lower black disk, so that in [Figure 10](#) the intersection of the balance loci is actually within the large black disk.

4. Grey-level ramps in the background. If the disks are placed on a background that shows a linear ramp, from lighter grey to darker grey, as in [Figure 11](#), then that has important implications for the Arnheim–Ross model. In effect, it is only possible to balance around one of the axes, there being no balance points for the other three axes. Thus, the vertical ramp in [Figure 11a](#) and [Figure 11c](#) means that there can be symmetry only around the vertical axis, and the horizontal ramp in [Figure 11b](#) and [Figure 11d](#) means that there can be symmetry only around the horizontal axis. The horizontal ramps varied from grey levels of 0 to 234 across the entire image width of 1024 pixels, therefore by about 117 greyscale units across the crop window. Likewise, the vertical ramps varied from 0 to 234 across the entire 768 pixels of the image, and therefore about 117 greyscale units across the crop window. A subtle point about, say, images 11a and 11c, is that the green line is closer to the upper disk in 11a and to the lower disk in 11c. That occurs because two areas of lighter grey and darker grey have been replaced by two areas which are equal in their blackness, and hence the balance point shifts slightly towards the disk in the lighter part of the image. An equivalent effect can be seen in [Figures 11b and 11d](#).

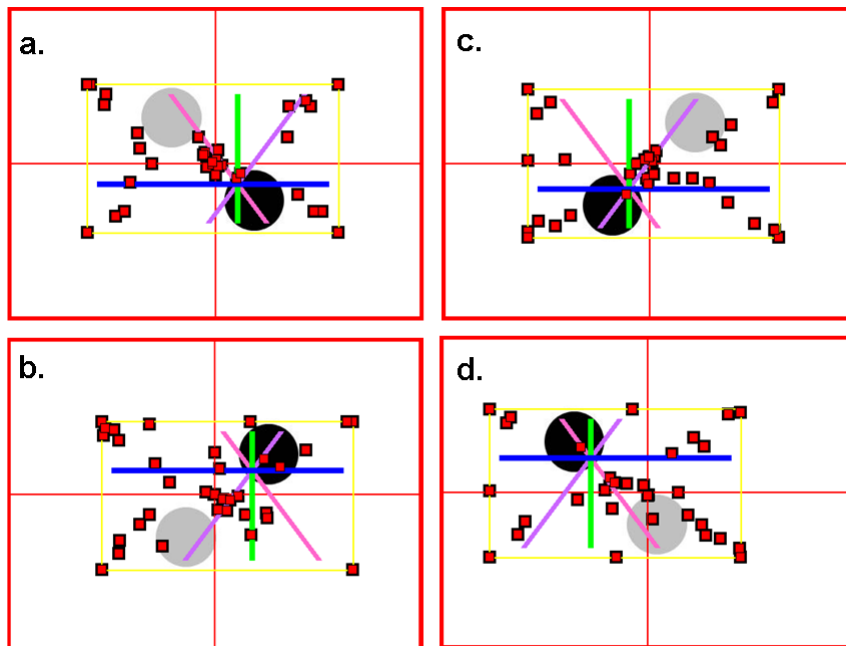


Figure 9. As for [Figure 8](#), but for the four images used to assess top-down and left-right anisotropy. Note that the crop window in these images, and those of [Figures 10 and 11](#), is rectangular in a 4:3 aspect ratio, rather than, as in [Figure 8](#), a square.

6.2.3 Participants. Thirty-six non-expert participants (12 male, 18 female) took part from the subject pool at University College London and were paid for their participation. The mean age was 28.75 (SD 9.18, range 19–61).

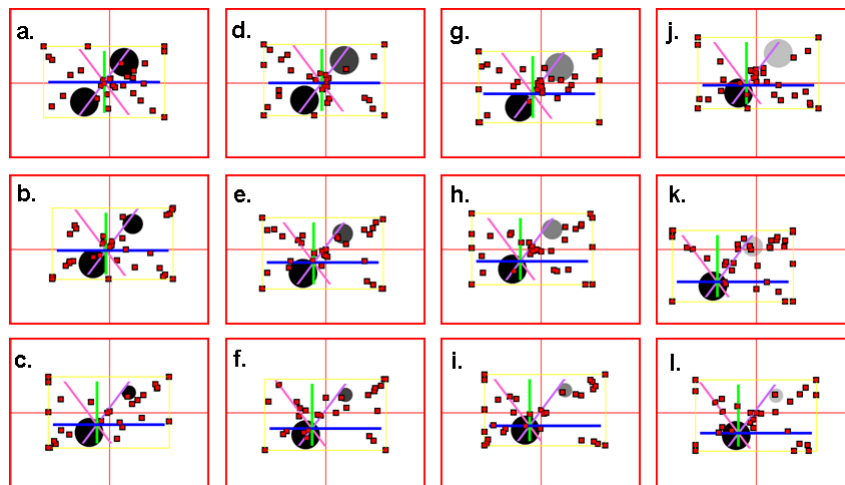


Figure 10. As for Figure 8, but for the 12 images assessing the effect of having a smaller or a lighter-grey disk in the top right-hand position.

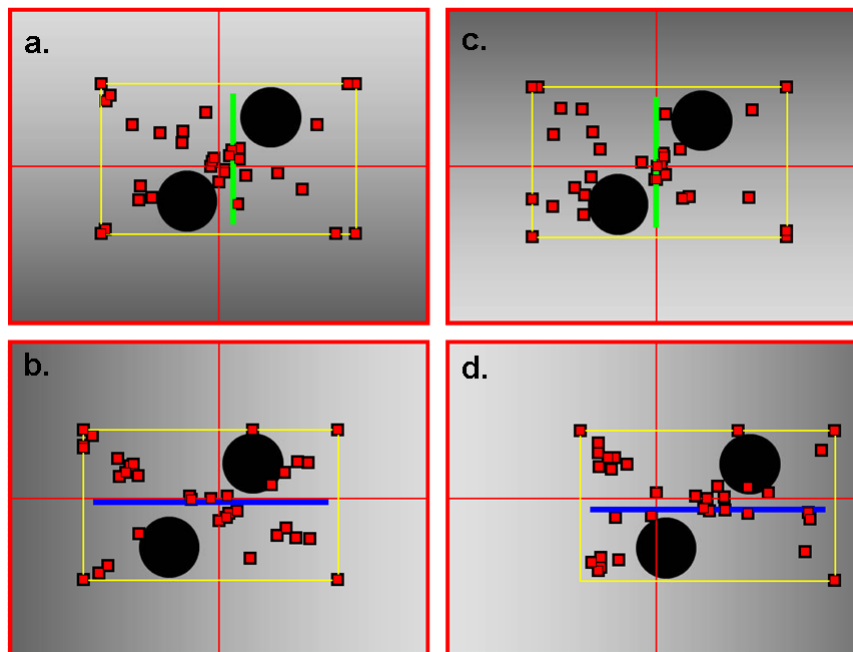


Figure 11. As for Figure 8, but assessing the effect of placing a grey ramp from top to bottom, bottom to top, left to right or right to left.

6.3 Study 5: Results

In Figures 8 to 11 the red box with the cross hairs at its centre shows the midpoint of the chosen crop window on each task (and hence, unlike Figure 6, the entire image space is not shown). An interesting feature of these data is that in 25% of cases (216/864) the participants have chosen points which lie on one of the limits of the positions the crop window can occupy, with 94 cases (10.9%) where the chosen point is in one of the corners. These limits were imposed because to allow the participants to crop further would be to crop the disks themselves, which would not be the situation that Arnheim had been considering. In order to avoid the descriptive and inferential statistics being distorted by these cases, a) in Figures 8 to 11 the average crop window is based only on those participants who did not place the crop window at an edge; and b) inferential statistics are based both on all participants and also

on participants who did not place the crop window at an edge. Statistical analyses used the Friedman rank test for related samples and are reported as 'p(all participants) [p(participants not on an edge)]'. If participants are, on average, placing their crop windows where the physicalist theory would predict, then the intersection of the midpoints of the average crop window (shown as the thin red cross-hair lines) should lie on one of the balance loci.

1. Square images similar to that used by Arnheim (Figure 8). Figure 8c is Arnheim's image, and on average the participants are placing the frame where Arnheim would have expected. That is also true for Figure 8e, with the average crop window lying centrally between the disks. Figures 8d and 8f are more problematic, as the midpoint of the average crop window is still midway between the disks, despite the fact that the physicalist model would expect it to have moved down and to the left, further from the pale grey disk and nearer to the black disk. The Friedman test found no significant differences in the placement of the frame between the conditions in the X direction ($p = .893$ [$p = .982$]) or the Y direction ($p = .768$ [$p = .653$]).
2. Anisotropy of left-right and up-down (Figure 9). The Friedman test found no significant difference in the placement of the average crop window (X: $p = .179$ [.119]; Y: $p = .172$ [.503]). As with the previous experiment, although the physicalist theory would expect the centre of the crop window to be placed nearer to the black disk than the grey disk, Figure 9 shows that in all four cases the middle of the crop window is placed about midway between the disks. In general, there is symmetry around the horizontal and vertical axes, and therefore given the absence of an overall effect, there is little point in searching further for anisotropy.
3. Variation in size and greyness of one of the disks (Figure 10). A physicalist theory would predict that the smaller and lighter the upper disk, then the more the CoM would be pulled down and to the left towards the black disk. The Friedman test found no significant difference in the placement of the average crop window in the X direction ($p = .701$ [.492]), but there was a significant effect in the Y direction ($p = .014$ [$p = .032$]). However, detailed scrutiny of Figure 10 suggests that the effects are not as would be predicted by the physicalist model (and if the model were predicting properly, then the horizontal axis of the average crop window should lie on the blue line). In 10c, 10e, 10f, 10g, 10h, 10i, 10k and 10l the crop window is closer to the grey disk than predicted, with no cases where it is further from the disk. In general the pattern is of the centre of the average crop window being at the midpoint between the disks, irrespective of their size or greyness.
4. Grey level ramps in the background (Figure 11). The grey level ramps effectively mean that there is only one axis on which a balance can be obtained, on the vertical axis for the vertical ramps (Figures 11a and 11c), and on the horizontal axis for the horizontal ramps (Figures 11b and 11d). There is no significant difference in the locations of the crop windows in the X direction ($p = .573$ [.176]) or the Y direction ($p = .342$ [.568]). There should also be a much reduced variance horizontally in Figures 11a and 11c and a reduced variance vertically in Figures 11b and 11d, but there seems to be no sign of that. Overall, the distribution of individual points looks very similar to those in Figures 8, 9, and 10, with a clustering of the majority of points at the centre and a few points on the diagonals in particular.

6.4 Study 5: Discussion

Considering only the stimuli with rectangular crop windows (Figures 9 to 11), Figure 12 shows the summed positions of the crop windows, expressed in terms of the mid-point of the two disks when they are of equal size (eg, Figure 9 and Figures 10a, d, g, j and Figure 11). There is a clear clustering at the centre, midway between the two disks; there are many points on

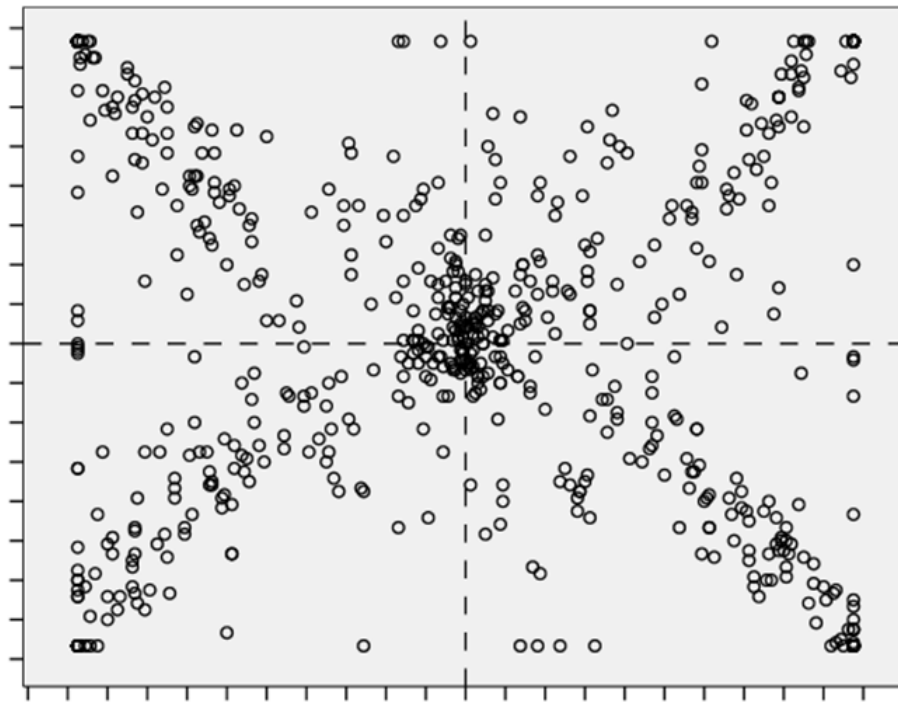


Figure 12. The results of all judgements in Study 5, aggregated across all stimuli where the crop window was rectangular and plotted relative to the midpoint of the two large black disks in [Figure 10a](#).

the diagonals; and in addition, there are points at the margins of the possible locations (and, particularly at the corners, these are often super-imposed).

Study 5 has directly replicated the process that Ross and then Arnheim recommended for finding the CoM of an image—moving a frame around the image until the image looks balanced. One of the stimuli was as close as possible to that described by Arnheim, and on average the participants did indeed place the frame where he would have expected, with the centre of the frame exactly between the two disks. However, in a range of other situations where the disks varied in size or greyness, or were placed in front of grey-level ramps, and where any model of balance would surely predict that the frame should be placed closer to the larger and heavier of the disks, the frame was still placed in almost exactly the same situation as with two identical black disks. Overall, the conclusion of Study 5 has to be that altering the size, greyness, or background of images has no effect at all upon their framing. That is a paradox, not least because to take, say, the stimuli of [Figures 10j, 10k, and 10l](#), there presumably has to come a point at which the disk becomes vanishingly small, or becomes as white as the background, when the frame would presumably, for Arnheim, have to be centred around the sole remaining disk.

7 General discussion and conclusions

This analysis of the Arnheim–Ross theory of balance in aesthetic images has been based on a literal, physicalist interpretation of Arnheim and Ross, one that uses the standard principles of statics, as understood by physicists, to investigate whether or not systems are balanced. Whether or not Arnheim and Ross meant to be taken in such strictly physical terms is not entirely clear, and it may be, as Cupchik (2007) argued, that the use of terms such as balance is primarily metaphorical. However, that does not really fit with the language and the level of description used by Arnheim and Ross. In particular, Arnheim (1971), as, for instance,

in his *Entropy and Art*, makes reference to a host of physicists and mathematicians, including Boltzmann, Bragg, Curie, Eddington, Kelvin, Lagrange, Mach, Newton, Planck, Schrödinger, Thomson, Tyndall, and Wiener, which suggests that he was thinking as a physicist, and meant his use of terms such as balance to be considered literally. Arnheim also understood the strengths (and the limitations) of mathematics in talking about art, quoting Leon Battista Alberti's comment at the beginning of his *Della Pittura* of 1435 that,

I shall primarily take from the mathematicians those things which will seem to me to be appropriate. ... But in all of my argument I wish it to be observed that I shall speak of these things not as a mathematician but as a painter (Alberti quoted by Arnheim 1966, p. ix).

If, despite this physical and mathematical underpinning, it is the case that Arnheim's theory is not to be interpreted in a physicalist way, it is not clear to us how it should be interpreted in order to obtain clear predictions for the various experimental situations we have described. And even if Arnheim intended only metaphor, more practically oriented artists have been more explicit, as was Ross, and as was also the case of Len A Doust, who in the 1920s and 1930s wrote a series of books on art (*The Doust Art Manuals*), in one of which (Doust 1934) it is said that, 'Composition is harmony, perfect balance; it is a sense of satisfaction in quantity' (p. 51, our emphasis). Doust continues by saying,

Let me define this problem of balance in a mechanical way. [Doust's] Figure 4c presents two unequal weights, which are balanced on a fulcrum in the centre of a picture; naturally, the bigger weight is nearer the fulcrum. ... This is the primary key of balance or composition. ... [G]ood composition depends upon a balance of unequal masses on the centre of the picture. ... One could write volumes on composition, and it would be but an elaboration of the simple fulcrum rule (pp. 545).

Likewise, and explicitly in the context of photography, Freeman (2007), immediately after a section on 'Gestalt perception', has a section on 'Balance', where there is a diagram of scales almost exactly as that in Doust, and it is said that,

When talking about the balance of forces in a picture, the usual analogies tend to be ones drawn from the physical world: gravity, levers, weights, and fulcrums. These are quite reasonable analogies to use, because the eye and mind have a real, objective response to balance that works in a very similar way to the laws of mechanics (p. 40, our emphasis).

That our physicalist interpretation of the Arnheim–Ross model, as defined above, does not explain our empirical findings seems clear enough. One possibility is that the essential form of the model is adequate, but that the parameterisation fails. It may be, for instance, that grey levels are not linearly related to perceived weight, or that distance is not weighted linearly to distance (as in equation (1) above); certainly, McManus et al (1985) found that the weight of squares is not always proportion to the square of their linear dimension but that individual subjects differ in the power relationship with different exponents. Although that is a possibility, it raises such a near infinity of possibilities, each potentially different for each individual perceiver, that it is not possible to take it further here. Whether or not it would save a physicalist interpretation of Arnheim is not clear, or perhaps it merely makes the theory unfalsifiable. Although testing only a subset of possible non-linear models, the analyses of Study 1 using equation (3) do strongly suggest that the position of the CoM is not particularly sensitive to the particular parameters of the model within a reasonable range.

In considering a physicalist interpretation of Arnheim, it should also be mentioned that Arnheim seems to be using a somewhat more complex theory when in *Art and Visual Perception* he discusses his Figure 5a and particularly 5b (our Figure 8a and 8b). In particular he says that the two disks 'attract each other', but that 'at a certain distance they repel each other because they are too close together' (p. 18). Arnheim also suggests that the two disks 'form a pair because of their closeness and their similarity in size and shape', and that 'as

members of a pair ... they are given equal value and function in the whole' (p. 18). Finally, Arnheim also discusses a series of attractive and repulsive forms within the space of a frame, which appear to be a repulsion from the frame itself and an attraction towards the centre (pp. 12 – 16). Such speculations would certainly result in a more dynamic, more sophisticated model than the simple physicalist account we have presented, far closer to Maxwellian fields, with objects and frame all having attractions and repulsions. Having said that, it is still difficult to see how such additional parameters and processes could account for the results seen in [Figure 10](#). The devil, as ever, is in the details.

Whether or not Arnheim was being physicalist there still remains a need to explain how composition in photographic and other images is determined. As Burgin (1982) said, the term 'composition' is used meaningfully in the visual arts, and that also applies in our studies, where participants make judgements which are consistent with one another. Clearly something is being done when participants compose images. Our studies have a number of positive conclusions, which should not be ignored. Our study of art photographs (Study 1) found that the centre of mass is indeed closer to one of the main axes than in control images (and in particular the external controls of randomly taken photographs). Although it seems unlikely that a physicalist account of balance can explain that result, some process must be underlying the difference, which yet remains to be found. Likewise, although the gamma ramp experiment (Study 2) failed to find any support for the balance point being on an axis, it did show clearly that in a majority of cases participants, and particularly experts, preferred the original image from the manipulated image which was only subtly different. Given that the image content, and especially its symbolic meaning, was effectively identical in the two images, the implication is that somehow formal, compositional properties of the images did indeed account for the preference. The cropping experiments likewise found that images cropped by participants had their centre of mass closer to an axis than did randomly cropped images (Study 3), which again supports the idea that some compositions are superior to others, although, again, the details of the spiders-web experiment suggest that the Arnheim–Ross model cannot explain the effect. Finally, in Study 5, which used the particular stimulus described by Arnheim, along with variants upon it, participants did not crop at random but showed highly systematic choices, even if those choices did not correspond to what the Arnheim–Ross theory predicts, and neither were they altered by obvious manipulations of size, greyness, or background. Overall, therefore, these findings, both positive and negative, do require explanation and suggest that the methodologies described here can be used to test such explanations.

Particular theoretical challenges for any theory of composition are posed by two of our results, which seem to pull in opposite directions:

1. In Study 5, altering the size or grey level of one of the two disks in Arnheim's original image, or using a vertical or horizontal grey-level ramp behind the disks, failed to alter the preferred framing position. That has to be a problem for any simple theory of balance which somehow integrates grey-levels across the entire image. A partial clue to the problem may have been suggested by Ross, who at the very end of his book states:

[W]hen any line or spot has a meaning, when there is any symbolism or representation in it, it may gain an indefinite force of attraction. This however is a force of attraction for the mind rather than the eye. ... The consideration of such attractions, suggestions, meanings, or significations does not belong to Pure Design but to Symbolism or to Representation (Ross 1907, p. 181).

The lack of any effect of the greyness of the circles, or of the background greyness, may result from the circles not being treated as merely a part of a gestalt of pixels but instead as objects so that it is the object status as circles, qua exemplars of circles (ie, symbols of

circles as well as just being circles), that overrides any mere consideration of the circles as being part of a total image. If so, then the challenge is to find images for which there is no such symbolic component (and it may be that purely abstract images may in some situations meet that requirement). Alternatively, it may be necessary to compare objects which differ in their symbolic force (so that although the attraction is large, it is not, in Ross's term, 'infinite', and hence can differ between objects within a frame). Finally, it might be felt that a partial solution to the problem would be provided were the images, in effect, to have been considered only in terms of their edges, so that grey levels were effectively removed, and in each case just two circles remained. That is so, but it seems difficult to understand how two differently sized circles should have equivalent force (either in a symbolic or a more physical sense); and even if that were the case, at the limit, as one circle becomes ever smaller until, like the Cheshire Cat, it disappears, the optimal balance at that moment would have to move so that the sole remaining circle would then be at the centre of the frame.

2. The precise antithesis of the problematic results of Study 5 is the findings of Study 2, the gamma-ramp experiment. These images were manipulated so that the gamma coefficient varied, in effect, in traditional photographic terms, the 'hardness' or 'softness' of the paper on which the image is printed, varied linearly across the horizontal and vertical dimensions of the image. That manipulation moved the centre of mass onto or off of a major axis. The physicalist interpretation of Arnheim's theory was not supported, as images with the centre of mass on an axis were not more preferred to those off an axis. However, and it is an important theoretical result, the original images were preferred over the manipulated images, despite the difference between the two barely being visible if the images were placed side-by-side (and hence the successive choice paradigm had to be adopted). The content of the original and manipulated images, in terms of what was being represented, was not altered, and therefore the images must surely have had near identical symbolic properties. The preferences for the originals must therefore be preferences which are largely independent of the symbolic meanings, and hence are underpinned by particular patterns of different grey levels across the image surface, despite those differences being hard to distinguish when the images are placed side-by-side. Finally, and it has to be important, the preference for originals was significantly greater in experts than non-experts.

In summary, any simple balance theory has severe problems because of Study 5, which seems to imply that it is symbolic values rather than any integral of grey levels which are important, whereas Study 2 seems to keep symbolic values constant, and yet original images are preferred to manipulated images, which must somehow depend on the integration of grey levels across the image, since that is all that has been manipulated.

An interesting finding, particularly in Study 5, is that there is a tendency always to centre the objects, irrespective of their actual weights or sizes. There is something similar here to the tendency, which has been repeatedly found, for eye-movements to concentrate towards the centre of images (Tseng et al 2009), even if they are blurred and of low resolution (Judd et al 2011), although there is a tendency for artists to show scanpaths covering more of an image (Vogt and Magnussen 2007). Since we have already referred to the work of Victor Burgin (1982), it is worth mentioning that he also considers the role of eye-movements in photographic composition, suggesting that as the eye explores an image then eventually it encounters the frame, at which point the risk is that the viewer will stop looking at the photograph and instead look outside the frame at the rest of the world. That risk,

may, however, be postponed by a variety of strategies which include 'compositional' devices for moving the eye from the framing edge. 'Good composition' may therefore be no more or less than a set of devices for ... retarding the recognition of the autonomy of the frame (Burgin 1982, p. 152).

Arnheim's theory, at least in our physicalist interpretation of it, is a Gestalt theory par excellence, the entire image contributing to its overall perception. In so far as many works of art are framed, and the entire image within the area is seen as contributing to the total aesthetic effect, then it makes sense for any theory of art images to be a Gestalt theory. Our physicalist interpretation of Arnheim and Ross is perhaps not the only type of theory which has such properties. Research in recent years has attempted to look at eye-movement data to find which aspects of visual scenes are likely to attract attention [and the work of Itti and Koch (2000) has concentrated on such issues, using low-level (non-object) information], and it has been extended and developed by Torralba et al (2006), using a Bayesian approach involving the likelihood of particular types of objects within specific scene types. In recent years, following the work of Mandelbrot on fractal geometries, a number of workers have argued that works of art are characterised by a particular relationship, in Fourier analytic terms, between the amplitude of particular harmonics and their frequency (the so-called $1/f$ relationship occurring when $\log(\text{amplitude})$ is a linear function of $\log(\text{frequency})$, with a slope of -1). Naturally occurring stimuli, as well as art works, typically show $1/f$ patterns [see, eg, Voss and Clarke (1975), Graham and Redies (2007), Redies (2007), Redies et al (2007), and Koch et al (2010), as well as studies of images by Jackson Pollock (Jones-Smith and Mathur 2006; Taylor et al 1999)]. Some experimental studies, such as Aks and Sprott (1996), but not all (McManus and Jeffs 1990), have found preferences for fractal-like patterns in experimental situations. It is possible that principles derived from fractal analyses might help in explaining some of the results found in our studies, and that is being investigated at present. Certainly any such theory would be fundamentally Gestalt in its approach, integrating information across the entire image. Finally, and we are grateful to a reviewer for pointing out to us, Gestalt theory ultimately considers objects within frames, and not patches of light. Wertheimer (1923) made the point very clearly when he said, in the opening of a paper,

I stand at the window and see a house, trees, sky.

Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have '327'? No. I have sky, house, and trees [translated by Ellis (1938, p. 71)].

Any adequate Gestalt theory of composition has surely, therefore, to take account of the meanings of objects qua objects as well as the patterns of light and shade in which they find themselves embedded.

References

- Aks D J, Sprott J C, 1996 "Quantifying aesthetic preference for chaotic patterns" *Empirical Studies of the Arts* **14** 1–16 ◀
- Arnheim R, 1954 *Art and Visual Perception: A Psychology of the Creative Eye* (Berkeley, CA: University of California) ◀
- Arnheim R, 1966 *Introduction to 'Creative Painting and Drawing' by Anthony Toney* (New York: Dover Publications) ◀
- Arnheim R, 1971 *Entropy and Art: An Essay on Disorder and Order* (Berkeley, CA: University of California Press) ◀
- Arnheim R, 1974 *Art and Visual Perception: A Psychology of the Creative Eye. The New Version.* (Berkeley, CA: University of California) ◀
- Arnheim R, 1982 *The Power of the Center: A Study of Composition in the Visual Arts* (Berkeley, CA: University of California Press) ◀
- Ash M G, 1998 *Gestalt Psychology in German Culture, 1890–1967: Holism and the Quest for Objectivity* (Cambridge, UK: Cambridge University Press) ◀
- Behrens R R, 1998 "Art, design, and Gestalt theory" *Leonardo* 299–303 ◀
- Behrens R R, 2002 "How form functions: On esthetics and gestalt theory" *Gestalt Theory* **24** 317–317 ◀

- Brainard D H, 1997 "The Psychophysics Toolbox" *Spatial Vision* **10** 443–443 doi:10.1163/156856897X00357 ◀
- Burgin V, 1982 "Looking at photographs" In V Burgin (Ed.), *Thinking Photography* (Basingstoke, Hampshire: Palgrave Macmillan), pp 142–153 ◀
- Cadik M, 2009 "Perceptual evaluation of color-to-grayscale image conversions" *Computer Graphics Forum* **27** 1745–1745 doi:10.1111/j.1467-8659.2008.01319.x ◀
- Cartier-Bresson H, 2006 *Scrap Book: Photographs 1932–1946* (London: Thames and Hudson) ◀
- Cupchik G C, 2007 "A critical reflection on Arnheim's Gestalt theory of aesthetics" *Psychology of Aesthetics, Creativity, and the Arts* **1** 16–16 doi:10.1037/1931-3896.1.1.16 ◀
- Daniels P C, 1933 "Discrimination of compositional balance at the pre-school level" *Psychological Monographs* **45** 1–1 doi:10.1037/h0093300 ◀
- Doust L A, 1934 *A Manual on Sketching Sea, Town and Country* (London: Frederick Warne) ◀
- Ellis W, 1938 *A Source Book of Gestalt Psychology* (London: Routledge and Kegan Paul) doi:10.1037/11496-000 ◀
- Fechner G T, 1876 *Vorschule der Aesthetik* (Leipzig: Breitkopf and Haertel) ◀
- Fechner G T, 1997 "Various attempts to establish a basic form of beauty: Experimental aesthetics, golden section, and square [Translated by Niemann, M, Qhehl, J and Hoege, H]" *Empirical Studies of the Arts* **15** 115–115 doi:10.2190/DJYK-98B8-63KR-KUDN ◀
- Freeman M, 2007 *The Photographer's Eye: Composition and Design for Better Digital Photos* (Oxford, UK: Ilex, Elsevier) ◀
- 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, 1979 *The Sense of Order* (London: Phaidon Press) ◀
- Gordon I E, 2004 *Theories of Visual Perception (3rd edition)* (London: Psychology Press) ◀
- Graham D J, Redies C, 2007 "Statistical regularities in art: Relations with visual coding and perception" *Vision Research* **50** 1503–1509 doi:10.1016/j.visres.2010.05.002 ◀
- Grundland M, Dodgson N A, 2007 "Decolorize: Fast, contrast enhancing, colour to greyscale conversion" *Pattern Recognition* **40** 2891–2896 doi:10.1016/j.patcog.2006.11.003 ◀
- Hopkinson C, 1937 "Denman Walso Ross (1853–1935)" *Proceedings of the American Academy of Arts and Sciences* **71** 543–546 ◀
- Itti L, Koch C, 2000 "A saliency-based search mechanism for overt and covert shifts of visual attention" *Vision Research* **40** 1489–1506 doi:10.1016/S0042-6989(99)00163-7 ◀
- Jacobson W E, 1933 "An experimental investigation of the basic aesthetic factors in costume design" *Psychological Monographs* **45** 147–184 doi:10.1037/h0093310 ◀
- Jasper C C, 1933 "The sensitivity of children of pre-school age to rhythm in graphic form" *Psychological Monographs* **45** 12–25 doi:10.1037/h0093301 ◀
- Jeffrey I, 2005 *The Photography Book* (London: Phaidon) ◀
- Johnson M, 1995 "The elements and principles of design: Written in finger Jello?" *Art Education* **48** 57–61 doi:10.2307/3193559 ◀
- Jones-Smith K, Mathur H, 2006 "Fractal Analysis: Revisiting Pollock's drip paintings" *Nature* **444** E9–E10 ◀
- Judd T, Durand F, Torralba A, 2011 "Fixations on low-resolution images" *Journal of Vision* **25** 14–14 doi:10.1167/11.4.14 ◀
- Koch M, Denzler J, Redies C, 2010 "1/f² characteristics and isotropy in the Fourier power spectra of visual art, cartoons, comics, Mangas, and different categories of photographs" *PLoS ONE* **5** e12268 ◀
- Lee A W P J, 2003 *Diane Arbus: Family Albums* (New Haven, CT: Yale University Press) ◀
- 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** 213–227 doi:10.2190/EKMD-YMN5-NJUG-34BK ◀
- Locher P, Gray S, Nodine C, 1996 "The structural framework of pictorial balance" *Perception* **25** 1419–1436 doi:10.1068/p251419 ◀
- Locher P, Stappers P-J, Overbeeke K, 1998 "The role of balance as an organizing principle underlying adults' compositional strategies for creating visual displays" *Acta Psychologica* **99** 141–161 doi:10.1016/S0001-6918(98)00008-0 ◀
- McManus I C, Edmondson D, Rodger J, 1985 "Balance in pictures" *British Journal of Psychology* **76** 311–324 doi:10.1111/j.2044-8295.1985.tb01955.x ◀
- McManus I C, Jeffs S, 1990 "Fractal dimensionality and preference for synthetic random dot patterns" In *Proceedings of the 11th International Congress on Empirical Aesthetics*, Budapest, Aug 22–25, pp 203–206 ◀

-
- McManus I C, Kitson C M, 1995 "Compositional geometry in pictures" *Empirical Studies of the Arts* **13** 73–94 ◀
- McManus I C, Zhou A, I Anson S, Waterfield L, Stöver K, Cook R, 2011 "The psychometrics of photographic cropping: the influence of colour, meaning, and expertise" *Perception* **40** 332–357 ◀
- Metzger W, 2006 *Laws of seeing* (Translated by L Spillmann, S Lehar, M Stromeyer, and M Wertheimer) (Cambridge, MA: MIT Press) ◀
- Palmer S E, 1991 "Goodness, gestalt, groups, and Garner: Local symmetry subgroups as a theory of figural goodness" In G Lockhead, J Pomerantz (Eds), *The Perception of Structure: Essays in Honor of Wendell R Garner* (Washington, DC: American Psychological Association), pp. 23–39 ◀
- Palmer S E, Guidi S, 2011 "Mapping the perceptual structure of rectangles through Goodness-of-Fit Ratings" Submitted ◀
- Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: Transforming numbers into movies" *Spatial Vision* **10** 437–442 doi:10.1163/156856897X00366 ◀
- Redies C, 2007 "A universal model of esthetic perception based on the sensory coding of natural stimuli" *Spatial Vision* **21** 97–117 doi:10.1163/156856807782753886 ◀
- Redies C, Hasenstein J, Denzler J, 2007 "Fractal-like image statistics in visual art: similarity to natural scenes" *Spatial Vision* **21** 137–148 doi:10.1163/156856807782753921 ◀
- Ross D W, 1901 "Design as a science" *Proceedings of the American Academy of Arts and Sciences* **36** 357–374 doi:10.2307/20021036 ◀
- Ross D W, 1907 *A Theory of Pure Design: Harmony, Balance, Rhythm* (Boston, MA: Houghton, Mifflin) ◀
- Taylor R P, Micolich A P, Jonas D, 1999 "Fractal analysis of Pollock's drip paintings" *Nature* **399** 422–422 doi:10.1038/20833 ◀
- Torralba A, Oliva A, Castelano M S, Henderson J M, 2006 "Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search" *Psychological Review* **113** 766–786 doi:10.1037/0033-295X.113.4.766 ◀
- Tseng P-H, Carmi R, Cameron I G M, Munoz D P, Itti L, 2009 "Quantifying center bias of observers in free viewing of dynamic natural scenes" *Journal of Vision* **9(7)** 1–16 ◀
- Verstegen I, 2007 "Rudolf Arnheim's contribution to Gestalt psychology" *Psychology of Aesthetics, Creativity, and the Arts* **1** 8–15 doi:10.1037/1931-3896.1.1.8 ◀
- Vogt S, Magnussen S, 2007 "Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen" *Perception* **36** 91–100 doi:10.1068/p5262 ◀
- Voss R F, Clarke J, 1975 "'1/f noise' in music and speech" *Nature* **258** 317–318 doi:10.1038/258317a0 ◀
- Wertheimer M, 1923 "Untersuchungen zur Lehre von der Gestalt" *Psychologische Forschung* **4** 301–350 doi:10.1007/BF00410640 ◀
- Wilson A, Chatterjee A, 2005 "The assessment of preference for balance: Introducing a new test" *Empirical Studies of the Arts* **23** 165–165 doi:10.2190/B1LR-MVF3-F36X-XR64 ◀