British Journal of Psychology (1985), 76, 311–324 Printed in Great Britain

Balance in pictures

I. C. McManus, D. Edmondson and J. Rodger

Three experiments are reported on the phenomenon of the subjective balance of pictures. Subjects were asked to place a fulcrum beneath a picture so that it looked balanced. In Expt 1 reproductions of works of art were used as stimuli. Pictures showed large differences in balance point; subjects showed smaller differences, unrelated to handedness or eye-dominance. Monochrome reproductions produced similar balance scores to coloured reproductions. Chopping a portion from one end of a picture showed that particular features are not the origin of the balance phenomenon, but rather the balance judgement incorporates an integration of information across the entire picture field. In Expt 2 abstract controlled stimuli were used; balance depended primarily upon the position of an object, and to a lesser degree on its size and colour. Experiment 3 used stimuli similar to those in Expt 2 and showed that subjects differed in the way that they extrapolated from side length to the mass of a rectangular object, that vertical position had no influence upon balance, and that the shape of a rectangle did relate to balance, the effect interacting with the degree of artificial perspective induced in the stimulus.

Paintings are often stated to be 'well balanced', or 'poorly balanced'. Is the use of the word 'balance' in such a context simply metaphorical, or may one utilize quasi-physical methods actually to measure pictorial balance? In this paper we describe the results of asking subjects to indicate, by positioning a fulcrum, the 'centre of gravity' of a picture or abstract stimulus.

Aestheticians are unanimous about the pictorial importance of balance. Poore (1903) claimed that 'of all the pictorial principles none compares in importance with unity or balance', and Tucker (1930) argued that 'balance is of the very greatest importance', and that 'picture balance is a necessity'. Arnheim (e.g. 1954) considered that picture balance was an 'indispensable factor in aesthetic composition'; he also considered that this balance could be measured, for 'every finite visual pattern has a fulcrum'. Feldman (1958) took the analysis further and differentiated symmetrical balance or mirror-image symmetry, which is 'often felt to embody excessive similarity and thus to be boring or facile', and a complex balance 'analogous to mechanical weight, with the centre of the picture as the fulcrum', visually 'heavy' items being balanced by visually 'lighter' items at a greater distance from the fulcrum. An explicit literal use of the concept of balance with respect to colours may be found in Schwarz (1968, p. 21).

The majority of the experimental literature on the phenomenon is relatively early and lacking in adequate statistical analysis. Pierce (1894), using simple geometrical figures, found that subjects placed a short line further from the fulcrum than a longer line, and that empty areas were 'lighter' than filled areas. He also noted 'great individual differences'. Puffer (1903) used a more sophisticated version of Pierce's experiment, subjects being asked to place one movable form in relation to several fixed forms, in order to produce the most aesthetically pleasing result. As well as the effect of solid and open figures she also found that interest was of importance, a postage stamp used on successive trials becoming progressively 'lighter'. Puffer also examined 1000 paintings 'of accepted merit' and claimed that almost all had a complex symmetry which resulted in a balance around the mid-point. Both Angier (1903) and Berlyne (1966) have produced evidence that confirms Puffer's suggestion that interesting objects or complex objects tend to be heavier. Angier, like Puffer, assumed that a well-balanced picture was balanced about the mid-point. Lund & Anastasi (1928) and Hubbell (1940), in both cases asking subjects to modify stimuli until they appeared balanced, also suggested that the central vertical axis was important to balance.

Langford (1936) adopted a different experimental approach, asking subjects to look at paintings while he recorded their eye-movements, and then asking them whether the fulcrum of the picture was to the left, or the right or was central. He found no relation between balance judgements and eye-movements or eye-dominance, and found good statistical reliability of the balance judgements. In order to produce unbalanced pictures he cut off one or other side of some of his pictures, but nevertheless he assumed implictly that in well-balanced pictures the fulcrum would be at the central vertical axis.

While most authors have felt that the central vertical axis of a picture was the implicit fulcrum, there is also a suggestion that Golden Section ratios are of importance (see McManus, 1980, for a review of the Golden Section in aesthetics). Berlyne (1969) asked subjects to use a cursor to best divide a painting into two parts. The results were not clustered around the mid-point but instead tended to divide the base in a Golden Section ratio. Whether such a task is related to the simple balance phenomenon is far from clear.

In addition to interest and complexity affecting visual weight, there are several reports of colours having their own individual weights. Bullough (1907) suggested that dark colours were heavier than light ones, his evidence being the apparent instability of painted triangles in which the top half was darker than the bottom half. Other investigators have asked subjects to rate the weight of differently coloured blocks (e.g. DeCamp, 1917; Warden & Flynn, 1926; Taylor, 1930; Payne, 1958), but have not obtained entirely consistent results. Monroe (1925) asked subjects to 'weigh' colours on a simulated balance. This same experiment was repeated by Pinkerton & Humphrey (1974), taking care to control for luminosity; the order of colours from heaviest to lightest was: red, blue, green, orange, yellow and white.

The following three experiments examined balance in both real paintings and in experimental stimuli.

Experiment 1

Method

Stimuli consisted of a series of postcard reproductions of works of art each being mounted on a piece of black card, and viewed from a distance of 130 cm, a typical postcard subtending a visual angle of 6·2 degrees. A white equilateral triangular fulcrum (side length 1 cm) could be positioned anywhere beneath the stimulus by the subject. The instructions, which were based on those of Langford (1936) were:

> In this experiment I would like you to judge the balance of the composition of a number of pictures. Imagine the picture you see to be supported only on the white triangle [= the fulcrum] and that the picture is free to revolve about this axis. A balanced picture gives the feeling that under these conditions it would hang level; one which is overbalanced to the right gives the feeling that the right side would sag, and conversely one overbalanced to the left would sag to the left. The estimated weight of the objects depicted is not the sole criterion of balance; for example a range of mountains on one side of the picture might be satisfactorily balanced by a small tree on the other side. Not just the materials portrayed but the impression of the entire picture is to be considered. Make your judgements by the way you feel about the balance of the picture, using any criteria which seem relevant.

Each subject saw 100 stimuli. The stimuli could be divided into three groups:

(1) *Unmodified pictures.* These 44 pictures comprised 12 portraits of single individuals [divided equally by sex and whether or not the left or right cheek was shown (McManus & Humphrey, 1973)], 26 landscapes and 8 'group portraits' (i.e. pictures in which more than one individual was portrayed).

(2) *Coloured vs. black and white pictures.* Forty pictures were presented either in black and white or in colour. The stimuli were divided into two series, A and B, each subject seeing just one series or the other. Each series contained 20 of the 40 stimuli in black and white, and the remainder in colour.

(3) *'Chopped' pictures.* Sixteen pictures were presented in modified form. Half had previously been judged as being balanced to the left and half to the right of centre. A proportion of between 9·8 and 29·9 per cent (mean = 17·6, SD = 4·7 per cent) of the total width was removed from the heavier side. Each subject saw half of the stimuli in their chopped form and half in their unchopped form.

Forty subjects took part in the experiment, being balanced for sex and handedness. The eye-dominance of the subjects was also assessed, but no attempt was made to balance for this factor across the other design variables. Handedness was classified by self-stated writing hand, and eyedness was assessed by sighting dominance through a cardboard tube.

Results

For each stimulus a 'balance score' was calculated, which took the value $2d/w$, where d was the distance of the fulcrum to the right of the centre of the picture, and w was the width of the picture as seen by the subject. The scores could therefore take a range between -1 and $+1$, a score of 0 indicating positioning at the exact centre, and $+1$ or -1 at the extreme edges of the picture. Golden Section division of a picture would give a balance score of $\pm$ 0·236.

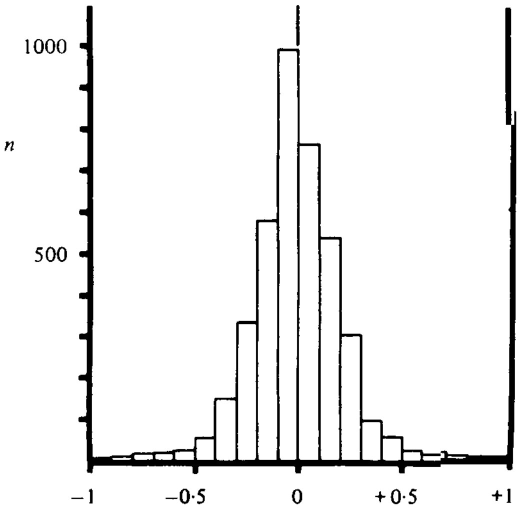

Figure 1. Experiment 1: The total distribution of balance scores.

Figure 1 shows the total distribution of balance scores produced during the experiment. It is clear that there is no tendency for the Golden Section division to predominate. Analysis of variance found significant differences between subjects ($F = 2{\cdot}97$, d.f. = 39, 3861, $P < 0{\cdot}001$; SD of subjects' means = 0·0268), and between pictures ($F = 8{\cdot}45$, d.f. = 99, 3861, $P < 0{\cdot}001$; SD of picture means = 0·0834). The overall mean of the balance scores in Fig. 1 is −0·0098 (i.e. slightly to the left of the centre). Using the subject × picture interaction as an error term, this value is significantly different from zero ($F = 10{\cdot}63$, d.f. = 1, 3861, $P < 0{\cdot}005$).

Using a regression approach analysis of variance none of the following—sex, handedness, eye-dominance, series of picture shown, nor any of their interactions—had a significant effect upon the mean score of subjects or upon the variance of the judgements of subjects. Principal factor analysis of the intercorrelations of the subjects' judgements suggested a single major dimension underlying subject differences.

The distribution of mean balance scores for individual pictures showed clustering around the centre with a wide distribution, a high proportion of works of art having balance points significantly removed from the centre. The Golden Section played no role at all, except in the negative sense that almost all of the means were between the two Golden Section division points. There were differences in variance of judgements between pictures (Bartlett–Box $F = 3{\cdot}35$, $P < 0{\cdot}001$), suggesting differences in the ease with which subjects may judge and agree on the balance point.

Coloured vs. black and white pictures. If colour is of importance in determining balance then we might expect to find that coloured pictures balance differently to black and white versions of the same pictures. The correlations between the mean judgements of the 40 pictures shown in coloured and monochrome versions was 0·839, and the regression equation was:

$$b_{\text{B\&W}} = 0{\cdot}891 \times b_{\text{colour}} + 0{\cdot}019.$$

The standard error of the regression coefficient was 0·094, and hence was not significantly different from unity ($t = 1{\cdot}16$, d.f. = 38, n.s.); there is thus no evidence that coloured pictures have consistently more extreme balance points than do black and white versions of the same pictures. A general test for *any* differences in balance between monochrome and coloured versions found no evidence for differences (chi-square = 28·73, d.f. = 40, n.s.).

Chopped vs. unchopped pictures. Sixteen pictures were presented complete or in a chopped version, a proportion having been removed from the previously determined 'heavier' side. For the chopped pictures two separate balance scores may be calculated: b_u, an uncorrected balance score which is estimated with respect to the picture *as presented* (i.e. so that it has a range of -1 to $+1$), and b_c, a corrected balance score which is based on the original picture width and centre (and hence it has a restricted range on the chopped side). For the unchopped original version we may calculate an original balance score, b_o. Positioning of the fulcrum at the same absolute position in the chopped and unchopped versions would result in $b_o = b_c \neq b_u$; and considering correlations, $r(b_o, b_c) > r(b_o, b_u)$. Conversely, positioning the fulcrum at a similar proportion of the distance along the base of each version, as presented, would result in $b_o = b_u \neq b_c$, and $r(b_o, b_u) > r(b_o, b_c)$. If the balance in a picture is primarily dependent upon the position of a crucial pictorial feature (as for instance might seem to be the case in Nash's *Totes Meer*, Fig. 2), then we can predict that removing a peripheral portion of the picture should have little influence upon the absolute balance point, i.e. $b_o = b_c$. Alternatively, if the sense of balance is based on an integration of the entire picture space, then there should be a close relationship between b_o and b_u.

The regression equations for the corrected and uncorrected balance score of the chopped picture on the balance score of the unchopped picture are:

$$b_u = \underset{(\text{SE} = 0{\cdot}291)}{1{\cdot}094} \times b_o + \underset{(\text{SE} = 0{\cdot}031),}{0{\cdot}0162}$$

$$b_c = \underset{(\text{SE} = 0{\cdot}184)}{-0{\cdot}249} \times b_o + \underset{(\text{SE} = 0{\cdot}019).}{0{\cdot}0073}$$

The correlation between b_u and b_o is 0·709 ($P < 0{\cdot}005$), while that between b_c and b_o is $-0{\cdot}3407$ (n.s.). The balance of a picture is therefore more dependent upon a global

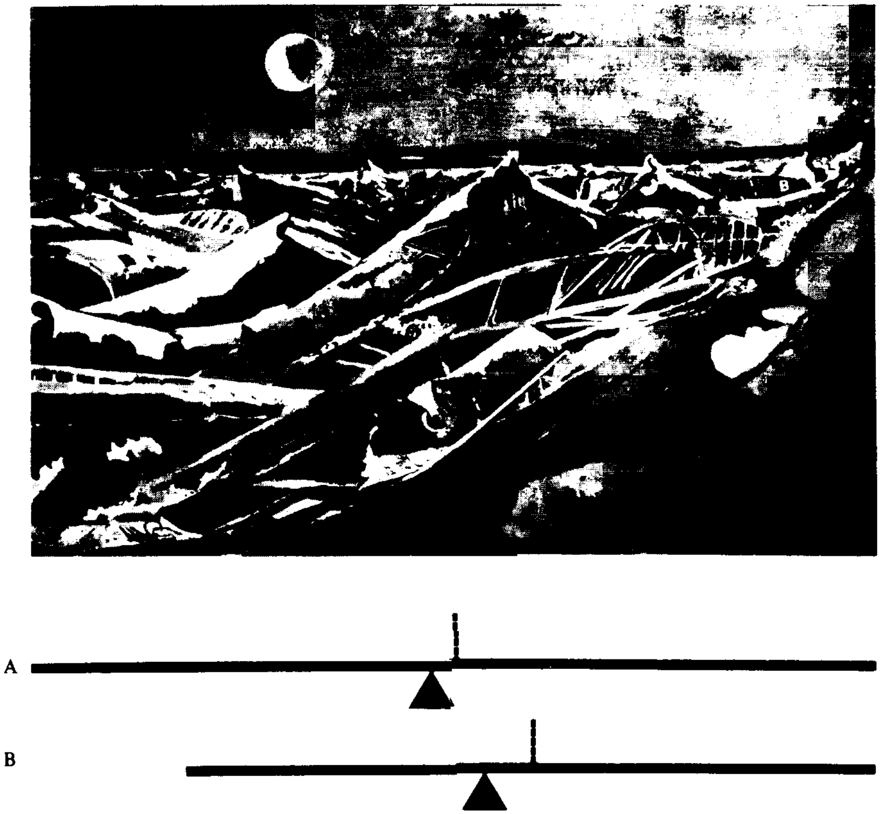

Figure 2. Experiment 1: Paul Nash's *Totes Meer*, painted 1940–41 (reproduced by permission of the Tate Gallery, London). Line A shows the mean fulcrum position when the picture was presented in its entirety; the aircraft wheel in the lower foreground appears to be particularly important in determining the balance point. Line B shows the mean fulcrum position when a modified version was presented in which the extreme 18·8 per cent had been removed from the left-hand side; the aircraft wheel is no longer of importance. Dotted lines indicate the centre of the original and of the chopped versions.

integration of the picture as a whole, than upon any individual element of it. This can be clearly seen in the case of Nash's *Totes Meer*, where the apparently critical aircraft wheel in the unchopped version becomes of no importance in the chopped version.

For right-chopped pictures, whose mean b_o was +0·079, the mean chopped scores were +0·169 (b_u) and −0·048 (b_c). For left-chopped pictures whose mean b_o was −0·072, the chopped balance scores were −0·129 (b_u) and +0·062 (b_c). In each case b_o is closer to b_u than to b_c, and indeed while b_u maintains the same sign as b_o, the sign of b_c is reversed. We may therefore reject any suggestion that an absolute pictorial feature is the primary determinant of the balance point. On average b_u values are more extreme than b_o values ($t = 3{\cdot}22$, d.f. $= 15$, $P < 0{\cdot}01$). The action of removing a portion of the picture has therefore been to shift the balance point further from the actual centre of the picture. This would be explicable if the picture contained a 'density peak' at some point between the midline and the extreme edge on the heavier side (i.e. some form of quadratic function); removal of a small portion from the heavier side, as long as it did not actually remove the peak itself, would increase the relative weight of the heavier half of the remaining picture, and therefore shift the balance point to a more extreme position.

Discussion

This experiment has demonstrated that subjects find the task reasonable and show good agreement, differences being relatively minimal and unrelated to sex, handedness or eye-dominance (in opposition to the suggestion of Valentine, 1962). Differences between pictures are far more substantial, and suggest that the conventional view that paintings 'of accepted merit' are balanced at the centre, is wrong. Neither is there evidence that the Golden Section plays a special role in picture balance. The effect of removing a portion from one side of a picture suggests that particular pictorial features are of little importance in determining the balance point; rather, the subject integrates across the entire picture space.

Experiment 2

In Expt 2 we wished to examine the role of size, position and colour of individual pictorial items in determining balance, and how items interacted in more complex abstract stimuli.

Method

Stimuli were generated by an ITT 2020 microcomputer, and displayed on a Panasonic video monitor. Each stimulus consisted of a rectangular frame enclosing a background of 60 fine white horizontal stripes against a black background. Superimposed on the background were either one (Expt 2*a*) or two (Expt 2*b*) squares, whose colour, size and position could be varied. Considering the total width of the rectangular frame to be 2 units, then the height of the picture was 1·10 units (i.e. an aspect ratio of 1:1.82). The display subtended a visual angle of 7·5 degrees at 125 cm.

Experiment 2a. In Expt 2*a* the single square was of side length 0·162, 0·309 or 0·456 units, and its centre was placed 0·245, 0·491 or 0·736 units to left or right of the midline. Three colours were used—red, blue and green—these being the primary colours produced by the monitor. The colour and contrast controls of the monitor were set to maximum and the brightness to minimum to ensure comparability among subjects. It was not possible to ensure that different colours were of equal luminance, or to control the brightness of the squares. All possible combinations of size, colour and position of square were presented, making 54 stimuli, which were given in a random order, the order being unique for each subject. Unbeknown to the subjects, after they had seen all 54 stimuli, a further 20 stimuli were presented which comprised the first 20 stimuli they had seen, to provide a measure of test–retest reliability.

Experiment 2b. Each stimulus contained two squares, each of which was of size 0·162 or 0·456 units, placed 0·245 or 0·736 units to the left or right of the midline, in either red or green. Each stimulus contained one square to the left of the midline and one to the right. All possible 64 combinations of squares were presented to each subject in a random order, the order being different for each subject. As with Expt 2*a*, after the 64 stimuli had been presented, the first 20 stimuli for that subject were re-presented in order to assess test–retest reliability.

Subjects used the potentiometer of a hand control to adjust the position of a white triangular fulcrum beneath the stimulus until they were satisfied that the stimulus was balanced; when they pressed a switch, the position of the fulcrum was recorded, and the display removed. Subjects worked at their own pace, although fairly rapid judgements were encouraged.

The 20 subjects consisted of second and third year undergraduates at Bedford College, most of whom were studying psychology. No attempt was made to balance for sex or laterality. No subject had studied art or art history beyond O-level, and none had been told about the purposes of the experiment. The instructions used were similar to those in Expt 1. All subjects were tested individually, firstly on Expt 2*a*, and then, after a short break, on Expt 2*b*. A typical session lasted between 30 and 45 min.

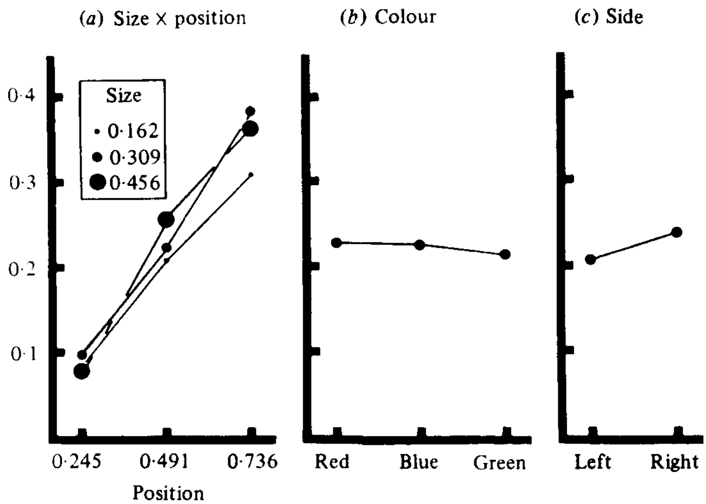

Figure 3. Experiment 2*a*: The ordinate is the modified balance score, a positive score indicating that the fulcrum was placed on the same side of the centre as the square: (*a*) interaction of size and position; (*b*) effect of colour; (*c*) effect of side.

Results

Balance scores for each judgement were computed as in Expt 1. The mean test–retest reliability of judgements in Expt 2*a* was 0·863 (SD = 0·095; range = 0·743–0·992) and in Expt 2*b* was 0·797 (SD = 0·208; range = 0·291–0·973).

Experiment 2*a*. A repeated measures analysis of variance was used to analyse the first 54 judgements of each subject, colour, size and position of the square being variates and the order of presentation of stimuli being a covariate. Size and position were partitioned *a priori* into polynomial effects; colour was partitioned *a priori* on the basis of a predicted order of colour weight into a contrast (red > blue > green). For convenience of analysis a modified balance score was used as the dependent variable; this value had a positive sign if the balance point was on the same side of the centre as the square in the stimulus, and had a negative sign if the balance point was on the opposite side of the centre. Using such a convention allows one to enter the side of presentation of the square into the analysis as a factor in its own right. The analysis of variance showed a significant effect of linear size ($F = 4{\cdot}62$, d.f. = 1, 37, $P < 0{\cdot}05$), linear position ($F = 240{\cdot}13$, d.f. = 1, 37, $P \ll 0{\cdot}001$), and their interaction ($F = 6{\cdot}81$, d.f. = 1, 75, $P < 0{\cdot}025$). There was no evidence for non-linear main effects of size or position, but there was significant evidence for a non-linear interaction between size and position ($F = 3{\cdot}22$, d.f. = 3, 75, $P < 0{\cdot}05$), which is shown in Fig. 3*a*. There were significant differences between colours in the predicted direction (Fig. 3*b*) ($F = 2{\cdot}97$, d.f. = 1, 37, $P < 0{\cdot}05$, one-tailed test), although it should be remembered that the colours were not equated for luminance. There was also a significant effect of side, the fulcrum being placed further from the centre if the square was on the right rather than on the left ($F = 4{\cdot}82$, d.f. = 1, 18, $P < 0{\cdot}05$) (Fig. 3*c*). No other interactions were significant at the 0·05 level of significance.

The covariate analysis showed significant regressions on order for the size stratum ($P < 0{\cdot}05$) and for the size × colour ($P < 0{\cdot}05$) and size × position ($P < 0{\cdot}01$) strata. Taken together these results suggest that the difference between large and small squares becomes slightly greater as the experiment proceeds.

Experiment 2*b*. For this experiment the dependent variable was the simple balance score, as used in Expt 1. The independent variables were colour, position and size, each of which had four levels according to the particular combination shown on the two sides; thus if there are two levels of a factor, A and B, then four display combinations are possible: AA, AB, BA and BB, the first symbol

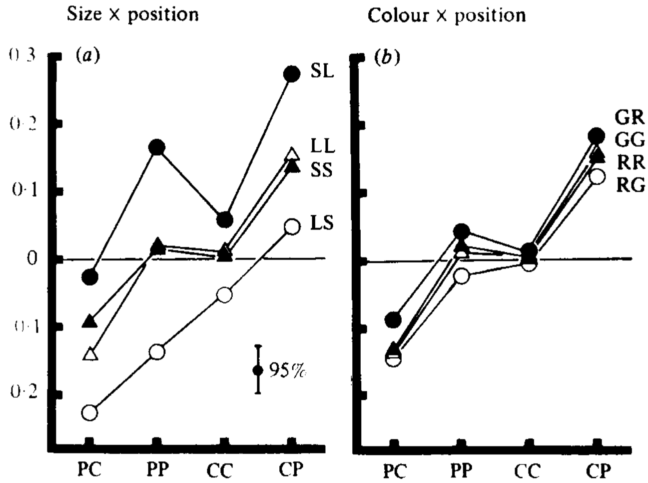

Figure 4. Experiment 2*b*: The ordinate is the simple balance score, positive values indicating a fulcrum to the right of centre, and negative to the left. The abscissa shows the four types of position used (P = peripheral; C = central), the first letter indicating the stimulus on the left: (*a*) the interaction of position with size (S = small; L = large); (*b*) the interaction of position with colour (R = red; G = green).

referring to the left-hand stimulus. Two *a priori* contrasts were considered for these factors. The more important contrast compared the AB condition with the BA condition; this should be the most sensitive test for subtle effects of size or colour, the two stimulus types, as it were, being placed in a set of scales. A second contrast compared the AA with the BB condition, and hence asked whether the different types of symmetrical conditions differed. For convenience the colours are denoted by R (red) and G (green), the sizes by L (large) and S (small) and the positions by C (central) and P (peripheral). The order of presentation of stimuli was used as a covariate.

There was a significant effect of position, peripheral objects shifting the balance point more than central objects, the CP vs. PC contrast being highly significant ($F = 161{\cdot}26$, d.f. = 1, 56, $P \ll 0{\cdot}001$) (see Fig. 4). There was also a significant effect of size shown in the LS vs. SL contrast ($F = 22{\cdot}03$, d.f. = 1, 56, $P < 0{\cdot}001$), large figures being heavier than small figures. Colour showed a significant effect in the RG vs. GR contrast ($F = 12{\cdot}16$, d.f. = 1, 56, $P < 0{\cdot}001$), red being heavier than green (Fig. 4*b*). Three interaction terms were significant: in the colour × position stratum there was a significant effect of RG vs. GR × CC vs. PP ($F = 6{\cdot}16$, d.f. = 1, 170, $P < 0{\cdot}025$) (see Fig. 4*b*); the differences between the asymmetric colour combinations RG and GR were greater when the squares were peripheral than when they were central. There were two significant effects in the position × size stratum (see Fig. 4*a*); there was a significant effect of CC vs. PP × LS vs. SL ($F = 35{\cdot}41$, d.f. = 1, 170, $P = 0{\cdot}001$); the difference between large and small squares was greater when the squares were both peripheral than when they were both central; there was also a small effect of PC vs. CP × LL vs. SL ($F = 4{\cdot}06$, d.f. = 1, 170, $P < 0{\cdot}05$); the difference between PC and CP conditions was shown better when both squares were large than when both were small (Fig. 4*a*). There were no significant covariate effects at any of the effect strata.

The mean overall balance score was $+0{\cdot}0177$ (i.e. slightly shifted to the right of centre), a value which is statistically significant when the between-subject term is used as its error ($F = 8{\cdot}08$, d.f. = 1, 19, $P < 0{\cdot}025$).

Discussion

It is clear from Expt 2 that the major determinant of balance in these simple stimuli is the position of an object in the field, and that the size and colour are of lesser importance, although they become relatively more important in stimuli which are almost symmetric. From Expt 2*a* there is a suggestion that the effect of size is a linear effect of the side length of the square, which contrasts with the physical situation in which weight should be proportional to the square of side length (if the objects are viewed as slices of equal thickness), or to the cube of side length (if the stimuli are felt to represent cubes). In most cases the interaction between position, size and colour is simple or additive, with one or two relatively minor exceptions.

Experiment 3

From Expt 2*a* there is a suggestion that the effect of size of stimulus is primarily linear, rather than quadratic. However, Expt 2*b* suggests that the design of Expt 2*a*, with only a single square in the display, is not very sensitive for detecting such effects. Experiment 3 was therefore designed to consider this question more carefully. In addition it was hoped to determine whether or not the balance of a stimulus was affected by the height–width ratio of the constituent rectangles, by the vertical position of the rectangles, and by the addition of a perspective cue in the background. Experiments 2*a* & 2*b* were completely balanced designs in which all possible stimulus combinations were given [in the case of Expt 2*b* the design being $(2 \times 2 \times 2 \times 2) \times (2 \times 2 \times 2 \times 2)$], each constituent square in a stimulus having a combination of four treatment levels. Clearly such a balanced design is impractical if one is interested in assessing a larger number of possible effects simultaneously in a reasonably sized experiment. The balanced design is also very tedious for subjects. A different approach was therefore used. The advantage of the balanced design is that it allows all possible interactions and polynomial effects to be assessed. However, in general most of these interactions are very unlikely *a priori*, and one is usually interested only in main effects and first-order interactions, with just linear and perhaps quadratic trends on each factor. A design was therefore used in which the four parameters of each component rectangle (height, H, width, W, and X and Y coordinates) were determined randomly within pre-specified ranges. The significance of the various effects was assessed by means of multiple regression, the design generally being approximately orthogonal due to the independence of the random numbers used in generating the stimuli.

Method

Stimuli were generated as in Expts 2*a* and 2*b*. Each stimulus contained two rectangles, one to the right of centre and one to the left. The height and width of the rectangles were uniformly distributed in the range 0·061–0·307 units. The centre of each rectangle was placed so that it occurred uniformly in the range 0·184–0·798 units to left or right of centre, and 0·184–0·920 units from the bottom of the picture frame. With these constraints it was not possible for any rectangle to touch the frame, to cross the midline or to abut or overlap the other rectangle. A sense of depth was induced in some stimuli by arranging the background lines of the display so that as they approached the top of the screen the lines became closer and closer, the illusion of a receding plane being produced. All rectangles were coloured red, and the background was of white lines against black.

Ten subjects took part in the experiment, all of whom were psychology undergraduates. Each subject saw 80 stimuli and made a judgement of balance as in Expt 2. The 80 stimuli for each subject were generated uniquely for that subject, and thus a total of 800 stimuli was used in the whole experiment. Each subject first saw a series of 40 stimuli in which no perspective effect was induced (i.e. in a manner comparable to Expt 2) and then saw 40 stimuli in which there was a perspective effect, the degree, P, varying on a scale from 0 to 1. The gap between the top of the frame and the first background line was set at 1/60th of the total vertical height of the frame (i.e. as in Expt 2). Succeeding gap widths down the display were calculated from the formula, $g_n = g_{n-1}(1+P/3)$. If $P = 0$ then the display was as in Expt 2. When $P = 1$ then there were only 16 lines in the background, with the lowest gap being six times as wide as the top one.

Results

Individual differences. A primary interest of Expt 3 was whether the size of the stimuli related to apparent weight in a linear, quadratic, cubic or higher-order manner, and whether subjects differed in that respect.

Consider a simple beam with two weights of mass M_L and M_R placed at distances D_R and D_L to left and right of the true centre. Let the fulcrum be placed at a point b' to the right of the centre. Then by the method of moments, at the balance point:

$$M_L(D_L+b') = M_R(D_R-b')$$

and hence

$$b' = \frac{D_R \cdot M_R + D_L \cdot M_L}{M_R+M_L}.$$

If each mass has a side of length L, and its mass is proportional to the pth power of L, i.e. $M_R = k \cdot L_R^p$ and $M_L = k \cdot L_L^p$, then

$$b' = \frac{d_R \cdot L_R^p - d_L \cdot L_L^p}{L_R^p + L_L^p}.$$

If two rectangles are being treated as *areas* in determining balance then p should equal 2, whereas if linear extent is the main determinant of mass then p should equal 1. Alternatively if the rectangles are being treated as three-dimensional objects then p should equal 3. If $p = 0$ this means that mass is of no importance in determining balance, and if p tends to infinity then the judgement is being made solely on the basis of the larger of the two stimuli.

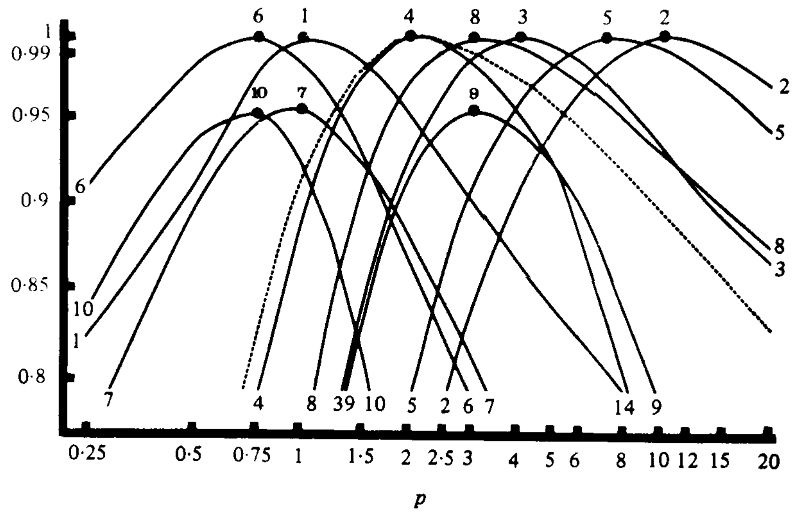

Figure 5. Experiment 3: Shows for each subject separately (numbered solid lines) and for all the subjects taken together (dashed line) the goodness of fit of the balance equation described in the text, in which weight is proportional to the pth power of geometric mean side length, values of p being shown on a logarithmic scale on the abscissa, and being calculated for values 0·25 (0·25) 5·0, 6·0 (1·0) 20·0. All curves are plotted relative to the best fit, which is indicated by a solid circle. The ordinate shows the proportional reduction in accounted variance relative to the best-fitting value of p. The curves for subjects 7, 9 and 10 have been shifted downwards by an arbitrary amount for clarity.

For each stimulus a predicted balance point, b', was calculated (L_R and L_L being calculated as the geometric mean of the rectangle side-lengths) and compared by a Pearsonian correlation with the actual balance point, b, for different values of p. Figure 5 shows the variation in goodness of fit of b' with b (expressed as the proportional reduction in accounted variance from the maximum accounted variance) for each subject. For subjects 1, 6, 7 and 10 the optimal value of p is 0·75 or 1·0, for subject 4 the optimal p is 2, for subjects 8 and 9, the optimal p is 3, and for subjects 3, 5 and 2 the optimal values of p are greater than 3, being 4, 7 and 10 respectively. Figure 5 also shows the same process for all 800 judgements from the 10 subjects combined; although $p = 2$ is the best fit, the spread is wider than for the individual subjects, and conceals individual differences. It is therefore clear that subjects differ in their perceptual judgement of mass, basing it on the 1st, 2nd, 3rd or higher powers of side length.

Effects of shape, position and perspective. Analysis of the effects of perspective and of size, shape and position of the component rectangles upon balance was by a multiple regression method, which is described in detail in the Appendix. The actual balance point was related to the predicted balance point from the simple physical model described earlier. The shape of the rectangle also had effects upon the balance, both as a main effect and in interaction with the horizontal and vertical positions of the rectangles, and with perspective. The main effect meant that the balance point was shifted further towards the taller and thinner of two rectangles than just a consideration of 'physical' factors would suggest (the shape of an object of course being unrelated to its effects on balance in a real physical system). In addition the effect of tall, thin stimuli was slightly less when the rectangles were placed relatively far from the midline, if the rectangles were higher in the picture field or if perspective was present, and was slightly greater if the taller rectangle was also the one furthest from the midline. The presence of perspective appeared to reduce the effect of tall, thin stimuli. There were no other effects of perspective cues upon balance, and neither were there any other effects of the vertical position of rectangles upon balance.

The constant term in the overall regression equation is $+0{\cdot}024$, suggesting a slight tendency for subjects to place the balance point to the right of centre.

Discussion

From Expt 3 it is clear that individual subjects differ in the manner in which they ascribe weight to rectangular stimuli, some using the linear, quadratic, cubic or higher-order functions of the geometric mean side length.The experiment confirms the importance of the mass of objects and their distance from the centre in determining balance, but suggests that subjects are not correctly integrating these two factors, at least as compared with the predictions of a simple physical model, even when individual differences in subjective mass are taken into account in the prediction equation. The induction of an artificial perspective had small effects upon the balance, although not in the expected manner, since from geometric illusions and size constancy one might expect that objects higher in the visual field would seem larger in the presence of perspective and thus should have a greater apparent weight; however, no evidence was found of an effect of vertical position in relation to perspective. The shape of stimuli did have an effect, although it was diminished by perspective, a result for which no obvious interpretation is apparent.

General discussion

From the experiments reported here it is clear that the balance of a picture, or of an abstract controlled stimulus, can be readily and reliably assessed by most subjects. There are differences between subjects, but in general the similarities are greater than the differences. The mechanical metaphor used by many theorists is true to a limited extent: objects further from the centre, or larger objects, cause a greater shift in the fulcrum from the centre. Nevertheless the 'weight' of objects was not primarily a response to physical characteristics, since subjects showed linear, quadratic, cubic or higher-order relations of subjective mass to linear extent, since position and size were not correctly integrated in judging balance, and since the shape of stimuli modified balance, the latter varying as a function of perspective. The experiments were therefore primarily measuring phenomenal or subjective weight rather than physical weight. That relatively large stimuli were not judged in the same way as small stimuli suggests that the frame of the overall stimulus might itself be exerting an effect upon the 'objects', perhaps in a Gestalt fashion, a result which emphasizes that in some sense the judgement of balance is an aesthetic judgement, the 'goodness' of the forms being assessed on the basis of the entire stimulus display. That subjects differed in their judgements of subjective weight may be explained in terms of differences in processing, in terms of differences in aesthetic judgement, or perhaps in terms of a failure of individual 'intuitive physics' (McCloskey, 1983) to correspond with veridical physics. Further experiments will be necessary to distinguish between these possibilities. Regrettably we do not have adequate evidence to begin to explain the individual differences shown in Fig. 5 in terms of characteristics of particular subjects. Indeed, as yet we are still not certain of the consistency of the phenomenon within individual subjects. Such questions should be of priority in further research.

From the study of paintings it is clear that in general paintings are not precisely balanced at the centre (and neither does the Golden Section play any significant role). It is of course possible, albeit unlikely, that artists differ from our subjects in their perception of balance, and that to their own eyes their pictures are balanced at a different point; such a prediction is empirically testable in living artists. Given the lack of precise balance in paintings it is questionable whether balance is as important as aestheticians have suggested, and indeed it might represent a *post hoc* rationalization for particular compositions.

Two previous studies of balance (Levy, 1976; McLaughlin *et al.*, 1983) have suggested that aesthetic preference for unbalanced pictures is different in right- and left-handers, and that asymmetric hemispheric processing is involved in this difference. By contrast, several other studies (Gorden & Gardner, 1974; Blount *et al.*, 1975) have suggested that subjects in general find it impossible to distinguish a picture from its mirror-image unless they have previously seen it, and hence balance is unlikely to be important in determining aesthetic response. The present experiments provide little comfort for a theory of aesthetics based on cerebral lateralization. In Expt 1 the mean balance point was shifted to the left of centre and in Expts 2*a*, 2*b* and 3 it was shifted to the right. The lack of consistency suggests that fundamental processes are unlikely to be of importance. More importantly, the lack of a difference in balance between right- and left-handers in Expt 1 provides no support for the position of Levy and McLaughlin *et al.*

Acknowledgements

Experiment 1 was carried out by J.R. as a BA project at the University of Reading, under the supervision of Prof. M. Coltheart. Experiment 2 was carried out as a BSc project by D.E. at Bedford College, University of London; we are grateful to Mr G. Rock for his assistance with computing in that study and in Expt 3. Prof. R. L. Reid read an earlier draft of the paper and we are grateful to him for pointing out an error in our reasoning about Expt 1.

References

Angier, R. P. (1903). The aesthetics of unequal division. *Psychological Review, Monograph Supplements*, **4** (1), 541–561.

Arnheim, R. (1954). *Art and Visual Perception*. Berkeley, CA: University of California Press.

Berlyne, D. E. (1966). Les mesures de la préférence esthétique. *Sciences de l'Art*, **3**, 9–22.

Berlyne, D. E. (1969). La section d'or et la composition pictorale occidentale et orientale. *Sciences de l'Art*, **6**, 1–5.

Blount, P., Holmes, J., Rodger, J., Coltheart, M. & McManus, C. (1975). On the ability to discriminate original from mirror-image reproductions of works of art. *Perception*, **4**, 385–389.

Bullough, E. (1907). On the apparent heaviness of colours. *British Journal of Psychology*, **2**, 111–152.

DeCamp, J. E. (1917). The influence of colour on apparent weight. *Journal of Experimental Psychology*, **2**, 347–370.

Feldman, F. B. (1958). *Varieties of Visual Experience: Art as Image and Idea*. New York: Abrams.
Gordon, I. E. & Gardner, C. (1974). Responses to altered pictures. *British Journal of Psychology*, **65**, 243–254.
Hubbell, M. B. (1940). Configurational properties considered 'good' by naïve subjects. *American Journal of Psychology*, **53**, 46–49.
Langford, R. C. (1936). Ocular behaviour and the principle of pictorial balance. *Journal of General Psychology*, **15**, 293–325.
Levy, J. (1976). Lateral dominance and aesthetic preference. *Neuropsychologia*, **14**, 431–445.
Lund, F. H. & Anastasi, A. (1928). An interpretation of aesthetic experience. *American Journal of Psychology*, **40**, 434–448.
McCloskey, M. (1983). Intuitive physics. *Scientific American*, **248**, April, 114–122.
McLaughlin, J. P., Dean, P. & Stanley, P. (1983). Aesthetic preference in dextrals and sinistrals. *Neuropsychologia*, **21**, 147–153.
McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, **71**, 505–524.
McManus, I. C. & Humphrey, N. K. (1973). Turning the left cheek. *Nature*, **243**, 271–272.
Monroe, M. (1925). The apparent weight of color and correlated phenomena. *American Journal of Psychology*, **36**, 192–206.
Payne, M. C. (1958). Apparent weight as a function of color. *American Journal of Psychology*, **71**, 725–730.
Pierce, R. E. (1894). Aesthetics of simple forms: Symmetry. *Psychological Review*, **1**, 483–495.
Pinkerton, E. & Humphrey, N. K. (1974). The apparent heaviness of colours. *Nature*, **250**, 164–165.
Poore, H. R. (1903). *Pictorial Composition and the Critical Judgement of Pictures*. New York: Baker & Taylor.
Puffer, E. D. (1903). Studies in symmetry. *Psychological Review, Monograph Supplements*, **4** (1), 467–539.
Schwarz, H. (1968). *Colour for the Artist*. London: Studio Vista.
Taylor, C. D. (1930). Visual perception versus visual plus kinesthetic perception in judging coloured weights. *Journal of General Psychology*, **4**. 229–246.
Tucker, A. (1930). Four talks on design: Balance. *The Arts*, **16**, 484–488.
Valentine, C. W. (1962). *The Experimental Psychology of Beauty*. London: Methuen.
Warden, C. J. & Flynn, E. L. (1926). The effect of colour on apparent size and weight. *American Journal of Psychology*, **37**, 398–401.

Received 22 November 1983; revised version received 8 November 1984

Requests for reprints should be addressed to I. C. McManus, Department of Psychiatry, St Mary's Hospital Medical School, London W2 or Department of Psychology, University College London, London WC1E 6BT, UK.

D. Edmondson is at the Applied Psychology Unit, Admiralty Research Establishment, Queen's Road, Teddington, Middlesex.

J. Rodger is at 9 Church Lane, Tasley, Bridgnorth WV16 4NW.

Appendix

Analysis of the results of Expt 3 requires an assessment of the effects upon balance of the 'weight' of the component rectangles (i.e. the *p*th power of geometric mean side length, *p* values being calculated separately for each subject), of their horizontal and vertical positions in the stimulus, of their shape, and of background perspective. Denote the masses of right and left components by M_R and M_L, the X and Y coordinates of their centres by X_R, Y_R, X_L and Y_L, their shape (calculated as the logarithm of height divided by width—see McManus, 1980), by S_R and S_L, and the degree of perspective in the background by P. For all effects except P we may consider the component which is different between the rectangles and the component which is common; only the former should have any direct effect upon balance. Similarly P alone should have no main effect upon balance, and will be considered as a common effect. We may thus calculate the difference components (denoted by 'd') as $d_X = X_R - X_L$; $d_Y = Y_R - Y_L$; $d_M = M_R - M_L$; $d_S = S_R - S_L$; and the common components (symbolized by 'c') as $c_X = X_R + X_L$; $c_Y = Y_R + Y_L$; $c_M = M_R + M_L$; $c_S = S_R + S_L$. There may be interactions between any of the c and d components with each other or with P. In view of the large number of possible interactions we have chosen, *a priori*, to look at those which seemed of possible importance. We considered all first- and second-order interactions between d components (i.e. $d_X \times d_Y$; $d_X \times d_M$; $d_X \times d_S$; $d_Y \times d_M$; $d_Y \times d_S$; $d_M \times d_S$; $d_X \times d_Y \times d_M$; $d_X \times d_Y \times d_S$; $d_X \times d_S \times d_M$; and $d_Y \times d_S \times d_M$), all first-order interactions between d components and c components (i.e. all combinations of one from d_X, d_Y, d_M, and d_S and one from c_X, c_Y, c_M, c_S and P); and second-order interactions between P and either two d components or one d component and a c component. In addition the predicted balance point based on simple physical theory was calculated for each stimulus (i.e. $b' = d_X + c_X \, . \, d_M/c_M$, which is equivalent to the formula given earlier).

Statistical testing was carried out by using a multiple regression method in which firstly for each effect being tested, E, a model was constructed in which b' and all interactions and main effects contained within E were added in to the model; on the second step the effect E was added and the increase in multiple squared correlation, R^2_E, found; on the final step nine variables were added which consisted of the nine interactions between subject means and E, the additional change in squared multiple correlation, $R^2_{S \times E}$, being found. The F statistic was calculated from the formula,

$$F(1, n-1) = \frac{(n-1) \, . \, R^2_E}{R^2_{S \times E}},$$

where n represents the total number of subjects in the analysis. For a completely orthogonal design this method of analysis is equivalent to a conventional repeated measures analysis of variance.

For the data of Expt 3 only eight of the 65 possible effects were significant, in addition to the effect of b', which was highly significant ($P \ll 0{\cdot}0001$). Three of the significant effects ($d_M \times c_X$, $P < 0{\cdot}05$; $d_M \times c_M$, $P < 0{\cdot}01$; and $d_X \times c_M$, $P < 0{\cdot}001$) were effects which are implicitly included in the predicted physical balance equation (given earlier), and suggest that subjects are not correctly integrating the separate information from the masses and from the distances from the centre. Of more interest is the significant effect of d_S ($P < 0{\cdot}05$) and of the interaction of differences in shape with perspective ($d_S \times P$, $P < 0{\cdot}01$), with distance from centre ($d_S \times c_X$, $P < 0{\cdot}05$), with difference in distance from the centre of the picture ($d_S \times d_X$, $P < 0{\cdot}05$), and with vertical distance from the bottom of the picture ($d_S \times c_Y$, $P < 0{\cdot}05$). The common component of d_S in these significant interactions suggests that they are not simply a Type I error due to repeated significance testing, but represent a genuine effect. These effects may be represented in a regression equation containing just these items and the predicted physical balance point:

$$\text{Balance} = 0{\cdot}392b' + 0{\cdot}128d_S - 0{\cdot}053\, d_S \times P - 0{\cdot}050d_S \times c_X + 0{\cdot}067d_S \times d_X - 0{\cdot}172d_S \times c_Y$$

which may be rearranged as:

$$\text{Balance} = 0{\cdot}391b' + d_S(0{\cdot}128 - 0{\cdot}053P - 0{\cdot}050c_X + 0{\cdot}067d_X + 0{\cdot}067c_Y).$$