Running head: Brightness-Weight Illusion

TO APPEAR IN EXPERIMENTAL PSYCHOLOGY

The Brightness-Weight Illusion: Darker Objects Look Heavier but Feel Lighter

Peter Walker,[1] Brian J. Francis,[2] Leanne Walker[1]

Lancaster University, UK

1 Department of Psychology

2 Centre for Applied Statistics

Abstract = 219 words

Word count (excluding references and tables) = 4998

Number of Tables = 2

Number of Figures = 1

Number of Footnotes = 2

Date of submission: 19th August 2009

Correspondence to:

Peter Walker, Department of Psychology, Lancaster University,

Lancaster LA1 4YF, UK.

e-mail: p.walker@lancaster.ac.uk.

tel: +44 (0) 1524 593163

fax: +44 (0) 1524 593744

Abstract

Bigger objects look heavier than smaller but otherwise identical objects. When hefted as well as seen, however, bigger objects feel lighter (the *size-weight illusion*), confirming that the association between visual size and weight has a perceptual component. Darker objects also look heavier than brighter but otherwise identical objects. It is uncertain, however, if this association also has a perceptual element, or if it simply reflects the fact that, in English at least, the same verbal label (*light*) is applied to both surface brightness and weight. To address this, we looked for a brightness equivalent of the size-weight illusion. Paired-comparison judgements of weight were obtained for balls differing only in color. Based on vision alone, darker objects were judged to be heavier. When the balls were hefted as well as seen, this association was reversed (i.e., a *brightness-weight illusion*), consistent with it having a perceptual component. To gauge the strength of the illusion (in grams), a white and a black ball (both 129 grams) were each compared against a set of mid-grey balls varying in weight. When the balls were hefted as well as seen, the white ball felt approximately 8 grams heavier than the black ball, a difference corresponding to 6.2% of their actual weight. Possible environmental origins of the association between surface lightness and weight are considered.

The Brightness-Weight Illusion: Darker Objects Look Heavier but Feel Lighter

The brightness of an object (or, more correctly, its surface lightness/reflectance) influences a person's judgement, reached by vision alone, of the object's weight, with decreasing levels of surface lightness suggesting increasing weight. This association emerges whether surface lightness varies in the context of achromatic colors (e.g., dark grey vs. light grey), or chromatic colors (e.g., dark blue vs. light red) (Alexander & Shansky, 1976; De Camp, 1917; Payne, 1958; 1961; Plack & Shick, 1976; Ryan, 1940; Warden & Flynn, 1926; Wright, 1962), though the association seems most pronounced in the context of achromatic colors (De Camp, 1917). Neither hue nor saturation (chroma) are consistently associated with visual judgements of object weight, though Alexander and Shansky (1976) claim that increasing levels of saturation are modestly associated with increasing weight.

We believe this association between surface lightness and weight has a perceptual component. Darker objects look heavier and are expected to feel heavier when lifted, in just the same way that bigger objects look heavier than smaller objects and are expected to feel heavier when lifted. There are other reasons, however, why people might say that darker objects are heavier. First, because the same verbal label in English is applied to both surface lightness and weight (i.e., *light*), the association could be linguistically mediated. Second, demand characteristics might be present when participants make visual judgements about the weights of objects that look identical except for their color. Intent on musing over the purpose of such an experiment, participants might well entertain the possibility that some aspect of the color of the objects is expected to influence their perceived weight. Given the presence of trials involving achromatic colors, participants might think that surface lightness is a key aspect, though it would remain unclear why they should assume that decreasing levels of surface lightness are associated with increasing weight.

How might the presence of a perceptual component in the association between surface lightness and weight be confirmed, while avoiding these alternative explanations? The strategy adopted in the experiment reported here was to determine what happens to the association when objects are hefted as well as seen. This strategy has been successful in relation to two other visual associations involving weight, namely, its associations with size (e.g., Flanagan & Beltzner, 2000; Murray et al., 1999) and material density (e.g., Ellis & Lederman, 1999; Harshfield & DeHardt, 1970; Wolfe, 1898).

When participants inspect, by vision alone, two objects that differ only in size, or only in apparent material density (e.g., brass versus styrofoam), the larger and denser objects are judged to be heavier. However, when participants also heft the objects, smaller and less dense objects are perceived to be heavier. This reversal of the associations (referred to as the *size-weight illusion* and the *material-weight illusion,* respectively) is believed to arise because the feedback from hefting fails to confirm *perceptual* expectations based on vision alone (i.e., that the larger/denser object will feel heavier).[1] According to Ross (1969), for example, perceptual expectations regarding the weights of objects guide preparations for lifting (e.g., grip force) and lifting itself (e.g., lift force). When objects are lighter than expected they are lifted more easily and quickly (see, for example, Davis & Roberts, 1976; Gordon et al., 1991). The ease and speed of lifting is then (mis-)interpreted as lightness in weight. Whatever the mechanism, if the association between surface lightness and weight has a perceptual component akin to the component contributing to the size-weight and density-weight associations, then hefting should reverse this association also.

Neither alternative explanation for the effect of surface lightness on judged weight predicts that hefting will reverse their association. First, allowing hefting does not reverse the lexical overlap between surface lightness and weight. Second, when asked, people do not expect the association between surface lightness and weight to change (let alone reverse) when hefting is allowed. Thus, whatever demand characteristics might be present with vision alone, they will remain in place with hefting. The critical question addressed in the present study is, therefore, whether the association between surface lightness and weight is reversed with hefting.

Experiment

To the authors' knowledge, only two studies have explored the possibility of a brightness-weight illusion (De Camp, 1917; Taylor, 1930). Whereas De Camp provides some data indicating the illusion exists, though for achromatic colors only, Taylor fails to find any effect of surface lightness on perceived weight when hefting is allowed. We used the method of paired comparisons to look again for a brightness-weight illusion, taking advantage of the current availability of objects produced commercially to high standards, and of advanced statistical techniques for analysing paired comparisons.

Materials

Eight, 52.4 mm diameter snooker cue balls (Aramith Tournament Champion), made from phenolic resin, were paint-sprayed in different matt colors. Three of the colors were achromatic (i.e., black, grey, white), five were chromatic (i.e., blue, red, green, lilac, turquoise). The five chromatic colors were specially mixed to match Munsell colors that sampled the whole color space, and allowed the impact of surface lightness and hue to be separately assessed (cf. Munsell Book of Color, 1976). Saturation was held constant at 8 in Munsell notation. The red (7.5 R) and blue (7.5 B) were intended to have the same lightness value (3 in Munsell notation), intermediate between black and grey. The lilac (7.5 P) and turquoise (7.5 G) were intended to have the same lightness value (7 in Munsell notation), intermediate between white and grey. The green (7.5 Y) was intended to have a lightness value matching the grey (i.e., 5 in Munsell notation). Room lighting was provided by two 40W daylight fluorescent tubes. When a sheet of white paper was placed on the surface where the balls were to be positioned, a luminance value of 89 cd/m^2 was registered (using a Minolta LS-100 luminance meter placed at participants' viewpoint). When small sheets of white plastic that had been paint-sprayed in the same matt colors as the balls were placed flat on the same surface, the luminance values (in cd/m^2) registered were: white = 89, turquoise = 41, lilac = 38, grey = 29, green = 24, red = 9, blue = 9, black = 5. The weights of the eight balls ranged from 128.4 to 129.8 grams (M = 129.15 grams, SD = 0.475 grams). A range in weight of 1.4 grams corresponds to 1.08% of the average weight of the objects, which is well below the typical Weber fraction of 5% for lifted weights.

Design

Each ball was paired with every other ball to yield 28 pairings. In a block of trials, these pairings were presented in a different random order to each participant, with the left-right positioning of the members of each pair being determined randomly. Every participant completed two blocks of trials, with the left-right positioning of the members of each pair counterbalanced across blocks. For one block of trials, participants indicated which ball was heaviest within each pair, and for the other block of trials they indicated which was lightest in weight. The ordering of these complementary judgements was counterbalanced within each condition.

Procedure

Each participant provided weight judgements under one of three conditions. In the *vision only* condition, participants looked at the balls, but were not allowed to touch them. They were asked to imagine hefting the balls and judge which one would feel the heavier/lighter of the two. In the *vision+heft* condition, participants were instructed to heft each ball while continuing to look at it. In the *heft only* condition, a black curtain hid the balls from view. Participants were permitted to heft the balls in whatever sequence they wished, and as many times as they wished, using the thumb and first finger of their right hand.

In all conditions, a trial started with the experimenter placing two balls on the beechwood desk in front of the participant. The balls were placed on two small fabric washers, separated by 10 cm, positioned 27 cm from the front edge of the desk. The washers were hidden from view when the balls were in position.

Participants

Thirty nine undergraduate students at Lancaster University, with normal or corrected-to-normal vision, completed the experiment. Thirteen students were allocated to each condition.

Results and Analysis

All participants responded to all the paired comparisons given to them, and no participant responded that the two objects being compared were matched in weight.

Figure 1 presents, separately for each condition, the probability with which the darker ball was judged to be the heavier of the two being compared as a function of the difference in their surface lightness. The colors were ranked according to their surface lightness, with pairs of colors intended to match in this regard being assigned the same rank. A difference in surface lightness of 0 refers to pairs of balls whose colors were intended to match in surface lightness (i.e., lilac/turquoise, grey/green, red/blue). However, to allow participants' responses to be coded consistently across all levels of difference in surface lightness, reference was made to the small residual difference in surface lightness within each of these 0-difference pairings (for which it is important to note that the red and blue balls had luminance values of 8.9 and 8.7 cd/m^2, respectively). A difference in surface lightness of 1 refers to pairs of balls whose colors were adjacent in their surface lightness ranking (e.g., white-lilac, green-blue, red-black). Differences in surface lightness of 2, 3, and 4 refer to pairs of balls whose colors were separated by 1 (e.g., white-green), 2 (e.g., lilac-black), and 3 (i.e., white-black) other levels of surface lightness, respectively. It will be noted that a difference in surface lightness of 4 arises only when the white and black balls were being compared. The number of data points determining each probability value was 78, 312, 208, 104, and 26, for differences in surface lightness of 0 through 4, respectively.

Inspection of Figure 1 indicates three things. First, for the *vision only* and *vision+heft* conditions, the probability values associated with a difference in surface lightness of 0 reveal that, despite the marked difference in hue within a pair of balls, each ball was equally likely to be judged to be heaviest. It would seem, therefore, that in these two conditions hue had no influence on perceived weight. Second, as expected, there was no association between surface lightness and judged weight in the *heft only* condition. Third, a systematic association between surface lightness and weight was apparent in both the *vision only* and *vision+heft* conditions. However, whereas in the *vision only* condition increasing levels of surface lightness were associated with lightness in weight, in the *vision+heft* condition increasing levels of surface lightness were associated with heaviness in weight.

Statistical Modelling

To evaluate these impressions formally, a statistical modelling approach was adopted to test nested hypotheses regarding the equality in weight of different subsets of balls. The Bradley-Terry model (Bradley & Terry, 1952) is designed for paired comparisons data, and is adapted from early psychometric work by Thurstone (Thurstone, 1927). Because paired comparisons yield discrete data, the model is more appropriate statistically than ANOVA-type methods which assume normality.

The model assumes that each item (hereafter, ball), j, has a “worth” (here, perceived heaviness), w_{jc}, in each condition, c. For most applications of the model, it can be fit as a special case of a logistic regression model with no intercept term. We develop the model as follows. Looking at all pairs within the comparison of balls j to k in condition c, if we observe N_{jkc} comparisons where j is judged heavier than k, and N_{kjc} comparisons where k is judged heavier than j, then N_{jkc} is binomially distributed with probability p_{jkc} and sample size $n_{(jkc)}$ (where $n_{(jkc)} = N_{jkc}+N_{kjc}$). For any pair of balls j and k, the model assumes that the probability that ball j is perceived to be heavier than ball k for condition c is:

$$\text{Prob}\ (j \textit{ heavier than } k \textit{ in condition } c) = p_{jkc} = \frac{w_{jc}}{w_{jc} + w_{kc}}$$

giving

$$\text{logit}(p_{jkc}) = \gamma_{jc} - \gamma_{kc} \qquad \text{where } \gamma_{jc} = \log w_{jc}$$

For identifiability, the parameter γ_{Jc}, for the last color category, J, is set to zero. The worth parameters are constrained to sum to one:

$$\sum_{j=1}^{J} w_{jc} = 1, \quad \text{giving} \quad w_{jc} = \frac{\exp(\gamma_{jc})}{\sum_j \exp(\gamma_{jc})}$$

However, left-right positioning needs to be controlled for in the analysis, and Dittrich et al. (1998) suggest a method, involving an extension of the Bradley-Terry model, to include both participant and item effects within complex designs. This extended model was used and fit as a generalised linear model using GLIM (cf. Francis et al., 1994).

We assume that the worths do not vary across individuals (i.e., there is no individual-level variability in the data). This assumption can be tested using goodness-of-fit to determine if the most complex model fits. If it does, then there is no evidence of individual-level variation.

Interest centres on whether certain of the worths are equal. For example, for a particular condition c, it is useful to determine if the worths, w_{jc} (or equivalently, the values for γ_{jc}), for all chromatic colors can be set to be equal. To do this, two logistic regression models are fit, one estimating a separate parameter, γ_{jc}, for each color, and the second constraining the parameters for the chromatic colors to be equal. The deviance, D (i.e., minus twice the log-likelihood), from each model can be interpreted as a goodness-of-fit statistic when compared to a chi-squared distribution on q degrees of freedom. Moreover, the difference in deviance, ΔD, and the difference in the degrees of freedom, Δq, can be calculated. A test for no difference between the chromatic color worths compares ΔD to a chi-squared distribution with Δq degrees of freedom. In general, such a test can be applied to any two models that are nested, that is, where one model can be obtained by constraining parameters in the other to be equal.

An alternative method of determining the best model is to calculate the Akaike Information Criterion (AIC) for each model (cf. Lindsey & Jones, 1998), which is defined as $D+2p$, where p is the number of parameters being estimated in a model. The model with the lowest value of AIC is taken to be the best fitting model.

Table 1 gives the deviances, degrees of freedom, and AIC values for a series of statistical models that impose varying restrictions on the worths of the colors. The most complex model, Model 0, assumes a left-right position effect, and estimates a parameter for each color separately for left and right position. Model 1 estimates an effect for each color, ignoring left-right position. Further simplification is achieved by restricting parameters for certain colors to be equal. Thus, Model 2 allows different parameters for white, grey, and black, but equates all chromatic colors to be equal. Model 4 equates white, grey, and black, but groups the chromatic colors into three surface lightness bands: red+blue, green, and pink+turquoise. Model 5 extends model 4 by also allowing the colors of black, grey and white to differ. Model 6, in contrast, groups all colors into five surface lightness bands: black, red+blue, grey+green, pink+turquoise, and white. The simplest model is Model 7, with all colors equated to each other. Figure 2 gives a pictorial representation of all eight models, showing which models are nested within which other models.

We first note that Model 0, the most complex model, fits for all three experimental conditions, with all three goodness-of-fit p-values being not significant. As Model 0 fits, there is no evidence of individual-level variation, and no evidence that different individuals have different worth profiles. Therefore, the Bradley-Terry model used is adequate for our purposes. To test for a left-right position effect, Model 0 was compared to Model 1, with changes of deviance being referred to a chi-squared distribution. All changes in deviance are small and insignificant, confirming that there was no effect of left-right position: $\Delta D = 10.77$, $df = 7$, $p = .15$; $\Delta D = 7.54$, $df = 7$, $p = .38$; $\Delta D = 5.17$, $df = 7$, $p = .64$, for the *heft only*, *vision only*, and *vision+heft* conditions, respectively. This factor was ignored in subsequent models. The analysis proceeded to find a parsimonious statistical model, with as few parameters as possible, by equating colors whenever there was no evidence of difference, and by using AIC values. The outcome for each of the conditions is as follows:

Heft Only. For the *heft only* condition, Model 1 can be simplified directly to Model 7, $\Delta D = 5.18$, $df = 7$, $p = .64$. Model 7 fits the data ($p = .49$) and has the smallest AIC value. Thus, in the absence of vision, all colors have equal worths (i.e., are perceived to be equally heavy).

Vision Only. For the *vision only* condition, Model 1 can be simplified to Model 5, $\Delta D = 0.47$, $df = 2$, $p = .79$. Model 5 can be further simplified to Model 6, $\Delta D = 0.01$, $df = 1$, $p = .92$, but not to Model 2, $\Delta D = 55.59$, $df = 2$, $p < .001$. Model 6 fits the data ($p = .75$) and also has the smallest AIC value. Thus, the worths of the colors are equated in each of the five surface lightness bands.

Vision+Heft. For the *vision+heft* condition, Model 1 can be simplified to Model 5, $\Delta D = 2.19$, $df = 2$, $p = .29$. Model 5 can then be simplified to either Model 6, $\Delta D = 0.80$, $df = 1$, $p = .37$, or Model 2, $\Delta D = 3.48$, $df = 2$, $p = .18$. Both models fit the data according to the goodness-of-fit test ($p = .54$ and $p = .48$, respectively). Neither of these models can be simplified to Model 7, $\Delta D = 39.99$, $df = 4$, $p < .001$, and $\Delta D = 36.51$, $df = 3$, $p < .001$, for Models 6 and 2, respectively. Therefore, both models, each of which is nested within Model 5, are supported by the data. However, Model 6 is preferred over Model 2 because the AIC value is slightly smaller and the goodness-of-fit p-value is larger. Thus, this condition provides results similar to the *vision only* condition, with the worths of the colors fitting into the five surface lightness bands.

Table 2 presents the parameter estimates and derived worths from the final chosen models for each of the three conditions. For the *heft only* condition, the worths of all the colors under Model 7 are set to be equal. For the *vision only* condition under the chosen final model (i.e., Model 6), the worths of the colors rank in reverse order to their surface lightness, with the darkest colors having the highest worth (i.e., perceived weight). For the *vision+heft* condition under the final model (again Model 6), this order is reversed, with the brightest colors now having the highest worths, though this is less pronounced for the chromatic colors than for the achromatic colors.

Ratios of worths can also be interpreted as odds ratios. Thus, in comparing the worths of the black and white balls, the odds of a participant preferring as heavier the black ball over the white ball is w_{black}/w_{white}. For the *vision only* condition, the odds ratio is 8.16 (i.e., 0.302/0.037), reflecting over eight chances of preferring black for every one against. For the *heft+vision* condition, the odds ratio is 0.292 (i.e., 0.060/0.201), or about 3 to 1 against. For the *heft only* group, all the odds ratios are one by definition.

In an earlier study, in which lighter (42.4 grams) 4 cm. wooden cubes served as objects, the same pattern of results was obtained. In the *vision+heft* condition, Models 2 and 6 were both supported by the data, but reference to their associated *p* values favored the latter. The reversed association between surface lightness and weight was again less pronounced for the chromatic colors.

Assessing the strength of the brightness-weight illusion

Before discussing the outcome of the experiment, we report some supplementary observations made when the method of constant stimuli was employed to assess the strength of the brightness-weight illusion. The white ball and the black ball from the experiment were each compared against a set of grey balls whose individual weights had been carefully adjusted. The intention was to measure the extent to which, in grams, the white ball felt heavier than the black ball in the *vision+heft* condition.

Materials

Nine snooker cue balls were paint-sprayed in the white, black, and grey colors used in the experiment. The white and black ball each weighed 129 grams. The grey balls had their weights adjusted to create a set varying from 115 to 157 grams in steps of 7 grams (i.e., in steps equivalent to 5.4% of the weight of the white and black ball). Their weights were adjusted by drilling into their centres of gravity, inserting varying amounts of denser material, and then back-filling with material of the same density as the phenolic resin. The values selected for the weights of the grey balls were based on the results of a small simulation that compared three alternative weight sequences. The range and spread of weights producing the smallest overall standard error for the log-odds estimate under a range of effect sizes was chosen.

Design

The white and black ball were each paired with every grey ball, once to the left and once to the right. The 28 trials generated in this way were presented in a random order, determined afresh for each participant. All participants completed two blocks of 28 trials, indicating which ball was lighter in weight in one block, and which was heavier in the other block. An equal number of participants completed the two blocks in each order.

Procedure

The procedure was identical to that adopted in the *vision+heft* condition of the experiment.

Participants

Twenty undergraduate students at Lancaster University, with normal or corrected-to-normal vision, provided data. None had been involved in the experiment.

Results and Analysis

All participants responded to all the paired comparisons given to them, and no participant responded that the two objects being compared were matched in weight. Two separate logistic regressions were carried out, one modelling the probability of choosing the white ball as heavier in the white-grey comparisons, and one modelling the probability of choosing the black ball as heavier in the black-grey comparisons. The varying weight of the grey ball was used as a single explanatory covariate in the regression. Thus

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1\text{weight}$$

The quantity of interest in both cases is $W50$, the estimated weight of the grey ball that would give a 50-50 chance of choosing the white (or black) ball as heavier than the grey. In the medical literature this is known as the median effective dose, or $ED50$. This can be estimated from the logistic regression by calculating $-b_0/b_1$. Likelihood-based confidence intervals (*CI*) can be calculated using relative likelihoods (see Aitkin et al., 2009, *pp.* 204-5, for further details).

The estimate of $W50$ for the white ball was 130.90 grams (*CI:* 129.71, 133.07), and for the black ball was 122.93 grams (*CI:* 121.54, 125.24). Thus, a grey ball with weight 130.90 grams is perceived to match a white ball of weight 129 grams, a difference of 1.90 grams. And a grey ball with weight 122.93 grams is perceived to match a black ball of weight 129 grams, a difference of 6.07 grams. The combined perceived weight difference between the black and white balls is thus 7.97 grams, equivalent to 6.2% of their actual weight, which is a little above the typical Weber fraction for lifted weights.

Discussion

In the experiment, there was no significant variation in the perceived weights of the objects in the *heft only* condition. The *vision only* condition confirmed that darker objects appear to be heavier in weight, for achromatic and chromatic colors alike. There was no effect of hue on perceived weight, with pairings matched for surface lightness (i.e., red-blue, green-grey, and lilac-turquoise) being judged to be similar in weight despite marked differences in hue. In the *vision+heft* condition, a brightness-weight illusion emerged, with darker objects now appearing to be lighter in weight. This reversed association was less pronounced for chromatic colors than for achromatic colors, in line with the observations of De Camp (1917).

The reversal of the relationship between surface lightness and perceived weight indicates that their association is unlikely to arise entirely from the presence of demand characteristics, or from the fact that in English surface lightness and weight share a verbal label (i.e., *light*). Of course, this lexical overlap could be a consequence, rather than a cause, of the association between surface lightness and weight. In any case, however, the same association has been observed with German language participants, for whom surface lightness and weight do not share a verbal label (Wright, 1962). It seems, therefore, that the reversal of the association between surface lightness and weight when hefting is allowed (the *brightness-weight* illusion) confirms the association has a perceptual aspect. As in the case of the size-weight and density-weight associations, it seems that perceptual expectations regarding weight are derived from the visual appearance of an object. Interestingly, for the visually perceived size of an object to influence how heavy it feels during hefting, the object has to remain in sight. If it is hidden from view just prior to and during hefting, the size-weight illusion does not occur (Masin & Crestoni, 1988). On the assumption that the same perceptual processes occur for both size and surface lightness, it is predicted that surface lightness will impact on perceived weight during hefting only if the object remains in sight so that its surface lightness can continually support the corresponding perceptual expectations.

The supplementary observations gave some indication of the strength of the impact surface lightness had on perceived weight when hefting was allowed. Though a 6.2% difference in the perceived weight of the white ball and the black ball might seem rather modest, it is important to note that it is greater than the Weber fraction typically observed with lifted weights (5%). Furthermore, for balls weighing 129 grams, 6.2% relates to an illusory difference of 8 grams. This is not trivial, especially given that surface lightness was competing with the ‘objective’ feedback provided by hefting. We expect surface lightness to have a much bigger impact on perceived weight when this is judged by vision alone (e.g., if people were to heft our modified grey balls and select one that feels as heavy as the white/black ball looks).

It is relatively clear where people will encounter an association between size and weight in the natural world and, therefore, how they might develop perceptual expectations linking visually perceived size to weight.[2] However, where might they experience an association between surface lightness and weight, so that corresponding expectations can be established? One can only speculate on this. It is possible that surface lightness and weight (i.e., material density) also are linked in the natural world. One referee of an earlier version of the present paper expressed the view that, in general, darker objects tend to be heavier objects, offering as an example the tendency for men to be darker and heavier than women. Certainly, most common materials (e.g., sand, soil, wood, and fabrics) do become darker and heavier when they are wet, and experience of this association alone might be sufficient to establish a general expectation that dark things tend also to be heavy things. Though a proper evaluation of the referee's conjecture requires a substantial study, we are able to report some very preliminary observations involving samples of rocks (i.e., pebbles) and samples of different types of wood. The surface lightness of each specimen was measured, and its density either measured or obtained from published sources. In the case of pebbles, there was no evidence of an association between surface lightness and weight. With regard to different types of wood, Edlin (1969) points out that they do not vary in density when trees are freshly cut. However, they do vary markedly after the wood has been thoroughly air-dried. Analysis of the 40 air-dried samples of wood provided in Edlin's book revealed a significant, though modest, association between surface lightness and weight, with darker timber tending to be heavier than lighter-coloured timber. It is pertinent to note, however, that the association became insignificant when one species of wood (ebony) was removed from the calculations. Clearly, further work is required to determine if surface lightness and weight are associated in the natural world in a way that allows corresponding expectations to be established.

References

Aitkin, M. A, Francis, B. J., Hinde, J. P. & Darnell, R. (2009) *Statistical Modelling in R*. Oxford University Press: Oxford.

Alexander, K. & Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. *Perception and Psychophysics, 19,* 72-74.

Bradley, R. A & Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika, 39,* 324-345.

Davis, C. M. & Roberts, W. (1976). Lifting movements in the size-weight illusion. *Perception & Psychophysics, 20,* 33-36.

De Camp, J. E. (1917). The influence of color on apparent weight: A preliminary study. *Journal of Experimental Psychology, 62,* 347-370.

Dittrich, R.J., Hatzinger, R. & Katzenbeisser, W. (1998) Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Applied Statistics, 47,* 511-525.

Edlin, H. L. (1969). *What wood is that? A manual of wood identification.* Stobart Davies: Ammanford, Carmarthenshire.

Ellis, R. R. & Lederman, S. J. (1998). The golf-ball illusion: Evidence for top-down processing in weight perception. *Perception, 27,* 193-201.

Ellis, R. R. & Lederman, S. J. (1999). The material weight illusion revisited*.* *Perception and Psychophysics,* 61, 1564-1576.

Flanagan, J. R. & Beltzner, M. A. (2000). Independence of perceptual and sensorimotor predictions in the size-weight illusion. *Nature Neuroscience, 3,*737-741.

Francis, B. J., Green, M. & Payne, C. J. (1994). *The GLIM4 System Manual.* Oxford University Press: Oxford.

Gordon, A. M., Forssberg, H., Johansson, R. S. & Westling, G. (1991). Visual size cues in the programming of manipulative forces during precision grip. *Experimental Brain Research, 83,* 477-482.

Gunlach, C. & Macoubray, C. (1931). The effect of color on apparent size. *American Journal of Psychology, 43,* 109-111.

Harshfield, S. P. & DeHardt, D. C. (1970). Weight judgement as a function of apparent density of objects. *Psychonomic Science, 20,* 365-366.

Lindsey, J.K. & Jones, B. (1998) Choosing among generalized linear models applied to medical data. *Statistics in Medicine, 17,* 59-68

Masin, C. M. & Crestoni, L. (1988). Experimental demonstration of the sensory basis of the size-weight illusion. *Perception & Psychophysics, 44,* 309-312.

Munsell Book of Color (1976). Macbeth: Baltimore, Maryland.

Murray, D. J., Ellis, R. R., Bandomir, C. A. & Ross, H. E. (1999). Charpentier (1891) on the size-weight illusion. *Perception and Psychophysics, 61,* 1681-1685.

Payne, M. C. (1958). Apparent weight as a function of color. *American Journal of Psychology, 71,* 725-730.

Payne, M. C. (1961). Apparent weight as a function of hue. *American Journal of Psychology, 74,* 104-105.

Plack, J. J. & Shick, J. (1976). Elementary school child's perception of weight as a function of color. *Journal of the Association for the Study of Perception, 11,* 16-30.

Robinson, E. J. (1954). The influence of photometric brightness on judgements of size. *American Journal of Psychology, 67,* 464-474.

Ross, H. E. (1969). When is weight not an illusion. *Quarterly Journal of Experimental Psychology, 21,* 346-355.

Ryan, T. A. (1940). Interrelations of the sensory systems in perception. *Psychological Bulletin, 37,* 659-698.

Taylor, C. D. (1930). Visual perception versus visual plus kinaesthetic perception in judging colored weights. *Journal of General Psychology, 4,* 229-249.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34,* 278-286

Wallis, W. A. (1935). The influence of color on apparent size. *Journal of General Psychology, 13,* 193-199.

Warden, C. J. & Flynn, E. L. (1926). The effect of color on apparent size and weight. *American Journal of Psychology, 37,* 398-401.

Wolfe, H. K. (1898). Some effects of size on judgements of weight. *Psychological Review,* 5, 25-54.

Wright, B. (1962). The influence of hue, lightness, and saturation on apparent warmth and weight. *American Journal of Psychology, 75,* 232-241.

Footnotes

1. Ellis and Lederman (1998) have shown that golfers' experience of the contrasting weights of practice and match balls (where practice balls are lighter in weight), a contrast that is not reflected in any visual cues of relevance to weight, gives rise to a reverse association (the *golf-ball illusion*) when the two ball types are arranged to be equally heavy and are hefted as well as seen. Thus, a ball with visual markings indicating it is a practice ball is perceived to be heavier than an identically weighted ball with visual markings indicating it is a match ball. Furthermore, the strength of the illusion is contingent on the extent to which a golfer has experienced the difference in weight of these two types of golf ball. This *golf-ball illusion* resembles the observation that after a particular object has been repeatedly judged to be heavy, because it has been compared against objects that are substantially lighter than it, the same object is perceived to be lighter than a new comparison object that is slightly lighter than it (cf. Murray, et al., 1999).

2. Objects with higher levels of surface lightness can appear to be bigger than equivalent objects with lower levels of surface lightness (e.g., Gunlach & Macoubray, 1931; Robinson, 1954; Wallis, 1935). A little reflection confirms that the effect of surface lightness on perceived weight cannot be mediated by its effect on perceived size. The nature of the association between surface lightness and perceived size is such that it would contradict the direct effect of surface lightness on perceived weight (e.g., because darker objects are judged to be smaller they would be judged to be lighter in weight, not heavier in weight).

Table 1 Deviances and number of parameters for different models

	Heft Only			Vision Only			Vision+Heft		
Models for color worths	Deviance (D) and df (q)	Goodness-of-fit *p*-value	AIC	Deviance (D) and *df* (q)	Goodness-of-fit *p*-value	AIC	Deviance (D) and *df* (q)	Goodness-of-fit *p*-value	AIC
0. All colors different, with separate estimates for each left-right position	40.09 42	0.55	68.09	36.73 42	0.70	64.73	41.86 42	0.48	69.86
1. All colors different	50.86 49	0.40	64.86	44.27 49	0.66	58.27	47.03 49	0.55	61.03
2. All chromatic equal, all achromatic different	53.74 53	0.45	59.74	100.33 53	<0.001	106.33	52.70 53	0.48	58.70
3. All achromatic equal, all chromatic different	52.30 51	0.42	62.30	134.12 51	<0.001	144.12	82.84 51	0.003	92.84
4. All achromatic equal, chromatic colors in three surface lightness bands	53.94 53	0.44	59.94	134.55 53	<0.001	140.55	84.98 53	0.003	90.98
5. All achromatic different, chromatic colors in three surface lightness bands	52.50 51	0.42	62.50	44.74 51	0.72	54.74	49.22 51	0.54	59.22
6. All colors in five surface lightness bands	54.97 52	0.36	62.97	44.75 52	0.75	**52.75**	50.02 52	0.54	**58.02**
7. All colors equal	55.68 56	0.49	**55.68**	187.09 56	<0.001	157.09	89.21 56	0.003	89.21

Note: The minimum value of AIC for each experimental condition is highlighted in bold.

Table 2 Color parameter estimates and derived worths for each condition (final models)

Color		Estimates of γ_{jc} (standard errors)			Derived worths, w_{jc}		
		Heft Only (Model 7)	Vision Only (Model 6)	Vision+Heft (Model 6)	Heft Only (Model 7)	Vision Only (Model 6)	Vision+Heft (Model 6)
Achromatic	Black	0	0	0	0.125	0.302	0.060
	Grey	0	-1.107 (0.199)	0.587 (0.180)	0.125	0.100	0.107
	White	0	-2.085 (0.241)	1.218 (0.212)	0.125	0.037	0.201
Chromatic	Red	0	-0.555 (0.199)	0.774 (0.181)	0.125	0.173	0.129
	Blue	0	-0.555 (0.199)	0.774 (0.181)	0.125	0.173	0.129
	Green	0	-1.107 (0.199)	0.587 (0.180)	0.125	0.100	0.107
	Lilac	0	-1.670 (0.196)	0.804 (0.181)	0.125	0.057	0.133
	Turquoise	0	-1.670 (0.196)	0.804 (0.181)	0.125	0.057	0.133

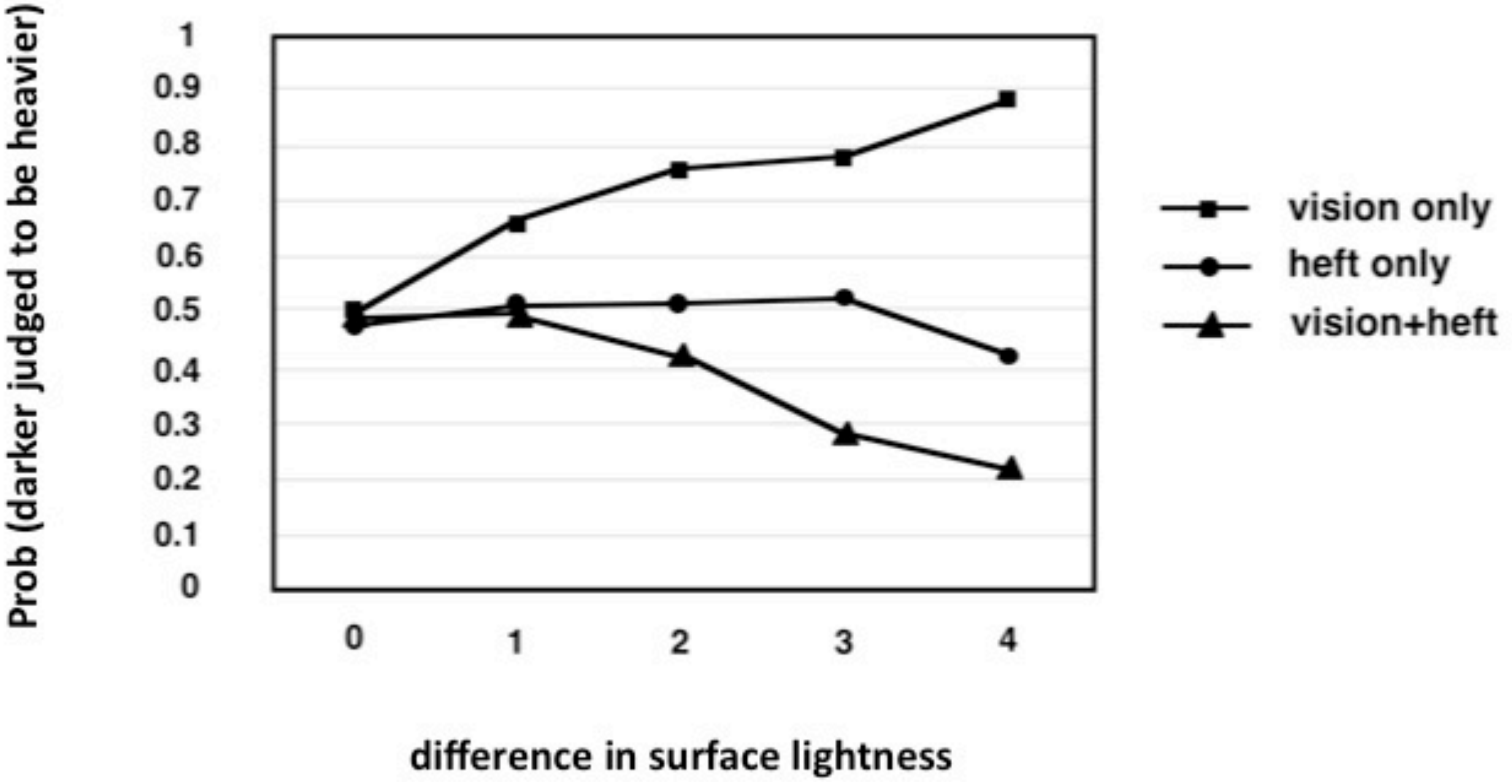

Figure 1. Plotted for each condition is the probability with which the darker ball of a pair being compared was judged to be heaviest in weight as a function of the difference in their surface lightness (0 = matching surface lightness, 4 = black versus white)

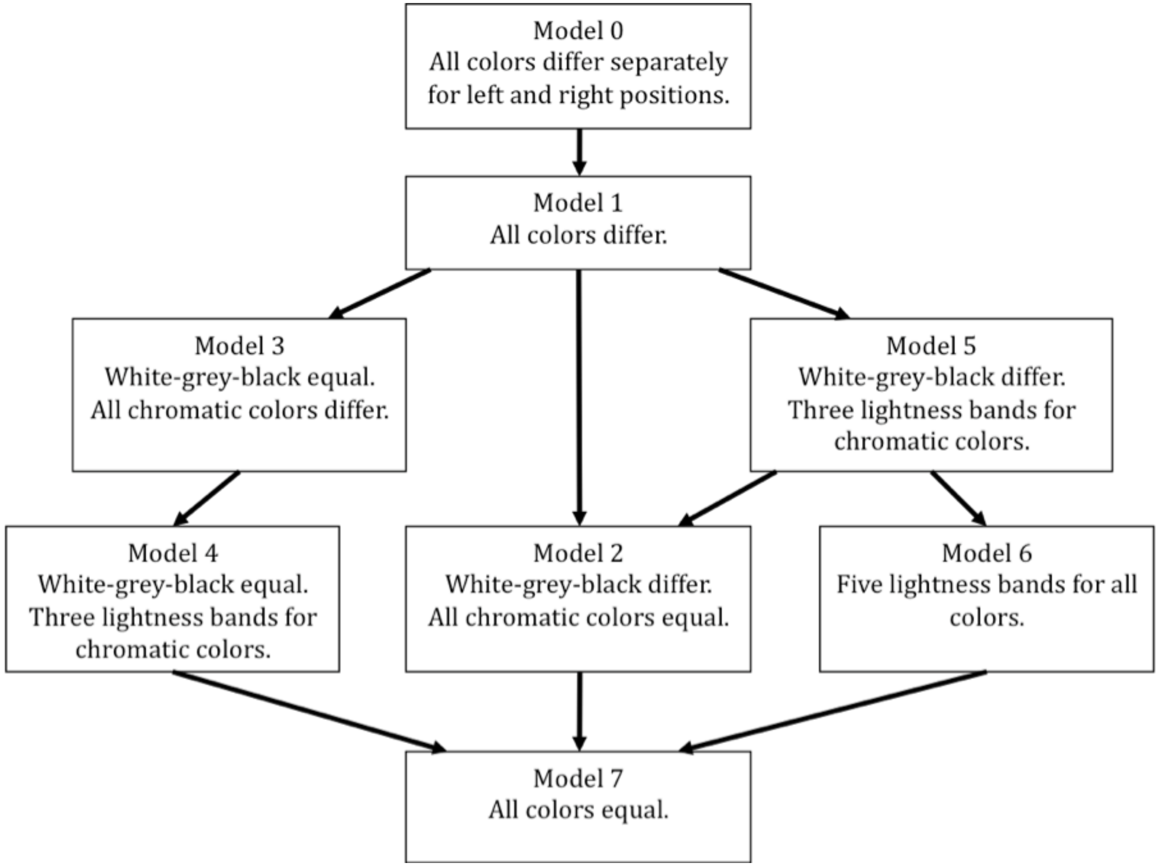

Figure 2. Model hierarchy for the various statistical models, showing which models are nested within which other models. Connected models can be tested through a likelihood ratio test.