Weight illusions explained by efficient coding based on correlated natural statistics

Paul M Bays

University of Cambridge, Department of Psychology, Cambridge, CB2 3EB, UK

Abstract

In our everyday experience, the sizes and weights of objects we encounter are strongly correlated. When objects are lifted, visual information about size can be combined with haptic feedback about weight, and a naive application of Bayes rule predicts that the perceived weight of larger objects should be exaggerated and smaller objects underestimated. Instead it is the smaller of two objects of equal weight that is perceived as heavier, a phenomenon termed the Size-Weight Illusion (SWI). Here we provide a normative explanation of the SWI based on principles of efficient coding, which dictate that stimulus properties should be encoded with a fidelity that depends on how frequently those properties are encountered in the natural environment. We show that the precision with which human observers estimate object weight varies as a function of both mass and volume in a manner consistent with the estimated joint distribution of those properties among everyday objects. We further show that participants' seemingly "anti-Bayesian" biases (the SWI) are predicted by Bayesian estimation when taking into account the gradient of discriminability induced by efficient encoding. The related Material-Weight Illusion (MWI) can also be accounted for on these principles, with surface material providing a visual cue for object density. This model framework could be used to predict perceptual biases in other properties that are correlated in the natural environment.

Keywords: efficient coding; Bayesian model; haptic feedback; vision; estimation

Introduction

The size-weight illusion (SWI), one of the strongest and most robust perceptual illusions, has been studied for over a century (Murray et al., 1999). Previous attempts to account for the illusion have appealed to violations of expectation (Ross, 1969), categorical representations of relative density (Peters et al., 2016) and optimality for long-distance throwing (Zhu & Bingham, 2011), among others. Here we show that efficient coding based on the empirical correlation between objects' masses and volumes provides a parsimonious explanation for the illusion based on the normative principle of efficient encoding.

According to the efficient coding hypothesis, sensory systems are optimized to transmit information about the natural environment (Attneave, 1954; Barlow, 1961). This can be achieved by distributing neural resources underlying encoding of sensory properties in accordance with the relative frequency with which those properties are encountered in the world (Ganguli & Simoncelli, 2014). This principle has previously been invoked to explain anisotropies in human judgments about visual orientation (Wei & Stocker, 2015). According to this account, discriminability is enhanced for cardinal orientations over obliques to reflect their greater prevalence in the natural environment. The resulting gradient of discriminability additionally causes orientation estimates to be biased away from the cardinals, a phenomenon termed likelihood

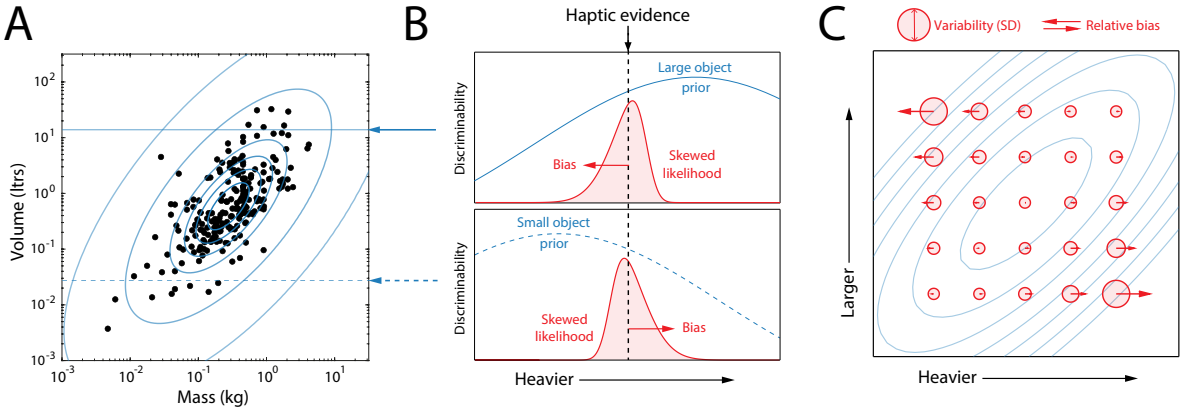


Figure 1: Predictions for bias and variability based on correlated environmental statistics. (A) Mass and volume of a large sample of liftable everyday objects (Peters et al., 2015) plotted on logarithmic axes. Blue contours show the best-fitting bivariate log-normal distribution (normal on these axes). (B) Illustrates estimation of mass for objects of larger (top) and smaller (bottom) volume. Blue arrows indicate corresponding volumes in (A). Discriminability (inverse of discrimination threshold or JND) varies in proportion to the conditional probability of object mass, given visual evidence of object volume (blue curves). For the same haptic feedback of object mass (black dashed line) the different gradients of discriminability cause likelihood functions (red) to skew in opposite directions. Red arrows show the relative directions of bias in mean posterior estimates of mass. (C) Predicted SD (circle diameters) and bias (arrows) in estimates of mass for objects with a range of true volumes and masses. Estimated weights of larger objects are underestimated relative to smaller objects of the same weight, as in the SWI.

repulsion. Subsequent work has generalized and refined the conditions under which this result holds (Morais & Pillow, 2018) and demonstrated that the predicted relationship between bias and the gradient of discrimination threshold, $b(x) \propto (D(x)^2)'$, is replicated across a wide range of stimulus variables (Wei & Stocker, 2017). However, while the theoretical model has been extended to multidimensional stimuli (Yerxa et al., 2020), empirical examples of these phenomena have so far been limited to single sensory variables. In this study we analyzed the case of properties whose occurrence is correlated in the natural environment: the size, weight and surface material of liftable objects.

Results

Correlated natural statistics

We began by considering the natural statistics of size and weight. Based on a large sample of liftable objects found in everyday environments (Peters et al., 2015), we observed that mass and volume have a joint density that is approximately bivariate log-normal with a strong positive correlation (bivariate normal on log axes, Fig. 1A; correlation, $\rho = .67$). As a consequence, visual evidence that a lifted object has larger or smaller volume would imply a conditional probability density for the object's mass that is log-normal with a peak at higher or lower mass, respectively. Efficient coding theory dictates that encoding resources should be allocated according to this natural frequency, inducing a gradient of discriminability for mass that reflects its conditional probability given the evidence for volume (blue curves in Fig. 1B, top and bottom).

For two objects with different volumes (one large, one small; Figure 1B, top and bottom) that invoke the same haptic feedback about mass when lifted (corresponding to arrow and dashed line), the differing gradients of discriminability result in likelihood functions for mass that are skewed in opposite directions. In other words, the internal representation is compatible

with a range of masses that deviates further in the direction of decreasing prior probability, because those masses are encoded with less precision. This produces a relative bias in point estimates of mass (e.g. the mean posterior estimate) such that the larger object is perceived as lighter than the small object, qualitatively matching the SWI. Fig. 1C illustrates how variability (circle diameter) and bias amplitude and direction (arrows) in mass estimates are expected to vary for objects of different mass and volume, assuming a correlated joint distribution (blue contours). Alternative encodings that also achieve the efficient coding objective are presented in Supplementary Information, and lead to qualitatively similar predictions, also consistent with the SWI (Fig. S1).

Fits to empirical data

Fig. 2 shows variability (black circles) and relative bias (black arrows) estimated from empirical observations of reported weight ratios of differently sized objects lifted by human participants (Peters et al., 2016). These results, which demonstrate a strong SWI, conform to the predicted pattern of Fig. 1C. Both estimation bias and variability depend on the combination of mass and volume in a manner qualitatively consistent with jointly correlated statistics (all $\mathrm{BF}_{incl} > 10^8$ indicating extreme evidence for effects of mass, volume and their interaction on bias and SD).

Red circles and arrows show mean predicted variability and bias, respectively, for the fitted

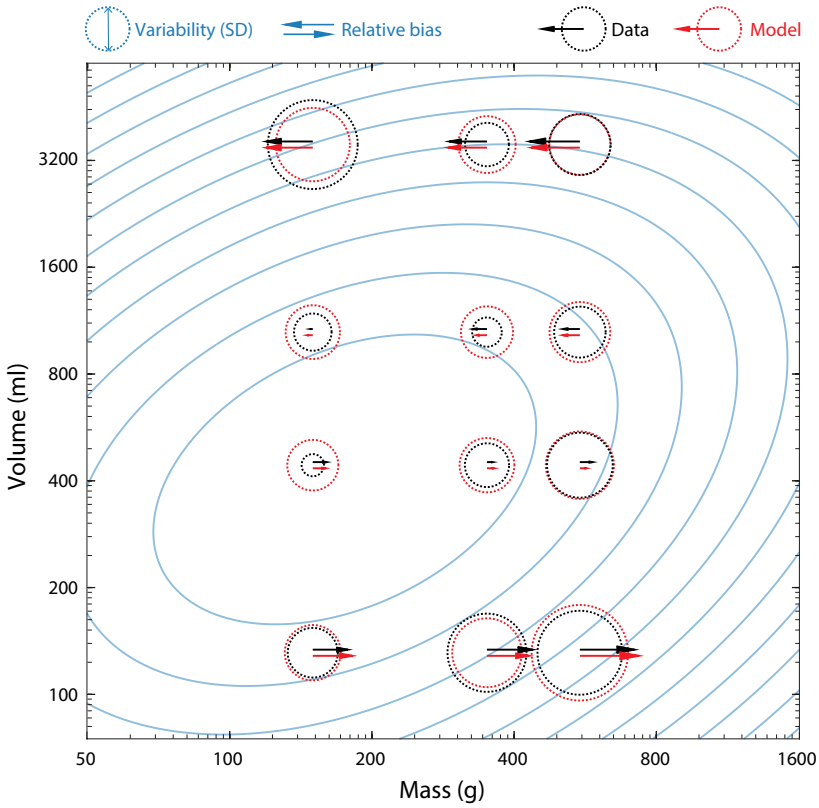


Figure 2: Summary data and fits. Black arrows and circles plot mean relative bias (arrows) and within-participant SD (circle diameters) of human observers in estimating mass, based on reported weight ratios in a task comparing pairs of lifted objects of equal weight and different volume (Peters et al., 2016). Red circles and arrows show corresponding mean predictions for the fitted efficient coding model with log-normal prior distribution (blue contours plot the prior density based on median fitted parameters).

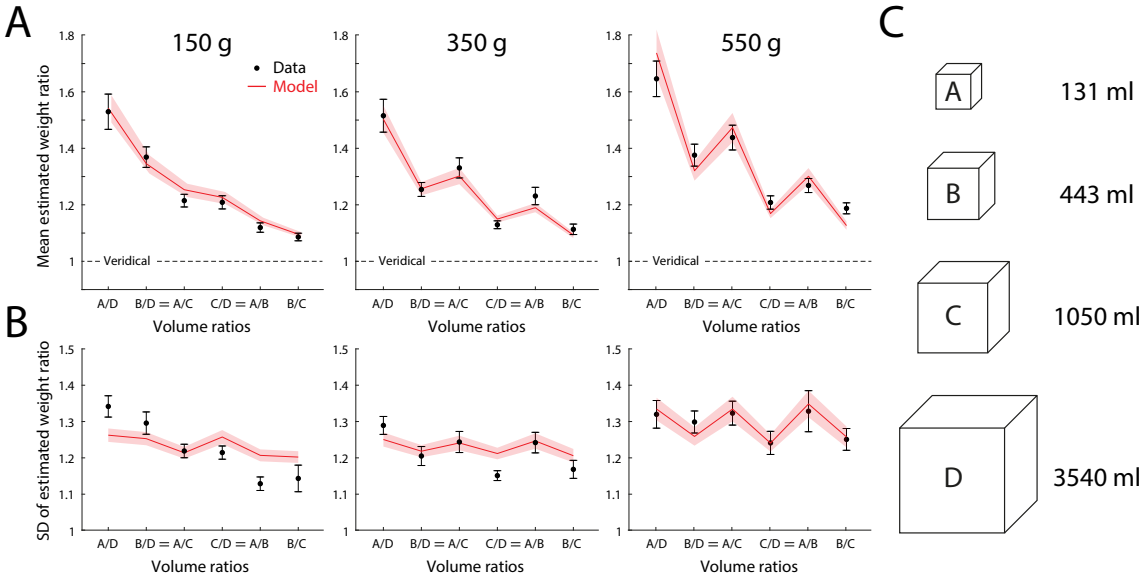


Figure 3: Estimated weight ratios and model fits. (A) Mean weight ratio reported by participants (black) for each pairing of lifted objects, with model predictions (red). Every pair of lifted objects had equal mass; estimated weight ratios greater than one indicate illusory overestimation of mass of the smaller object (the SWI). Comparisons within each panel are plotted in decreasing order of volume ratio (largest volume discrepancy on the left; equals signs (=) indicate different object pairs with equal volume ratio. Different panels show results for three sets of objects with different common masses. Error bars and shading indicate ±1 SEM. (B) Within-participant SD of weight ratios, plotted as in A. (C) Shape and volume of the objects.

efficient coding model (maximum likelihood parameters, median values: prior intercept, $c_m = 1.66$; prior slope, $\beta = 0.358$; prior variance, $s^2 = 11.2$; bias constant, $c_b = 0.0175$; s.d. constant, $c_\sigma = 0.00726$).

Fig. 3A (black circles) plots the mean estimated weight ratios for individual pairs of lifted objects from which the empirical estimates in Fig. 2 are calculated, along with predictions of the fitted efficient coding model (red lines). Despite each lifted pair having equal mass, the object of larger volume was on average estimated as heavier (weight ratio > 1) for every pair of objects in each mass condition, demonstrating the SWI. The strength of the SWI decreased as the ratio of the volumes decreased (left to right within each panel), and increased on average as mass of the objects increased (left-hand to right-hand panels). For pairs of objects with the same volume ratio (labelled with equals sign in Fig. 3), the SWI was in most cases stronger for the smaller pair (generating the zig-zag pattern in Fig. 3A middle and righthand panels). Each of these patterns was quantitatively reproduced by the efficient coding model (red lines), based on the nonlinear relationship between mass, volume and prior probability depicted in Fig. 2 (blue contours).

Fig. 3B (black circles) plots the within-participant s.d. of weight ratio estimates in the same format as Fig. 3A. Variability in estimation increased with object mass (lefthand to righthand panels), and for pairs with equal volume ratio, was in most cases greater for the smaller pair (zig-zag pattern in middle and righthand panels). The efficient coding model quantitatively captured the average SD within each mass condition, the increase in SD with mass, and also the pattern of variation in SD across pairs for the heaviest objects (550 g), although it underestimated the amplitude of variation in SD across object pairs for the 150 g (and to a lesser extent the 350 g) objects.

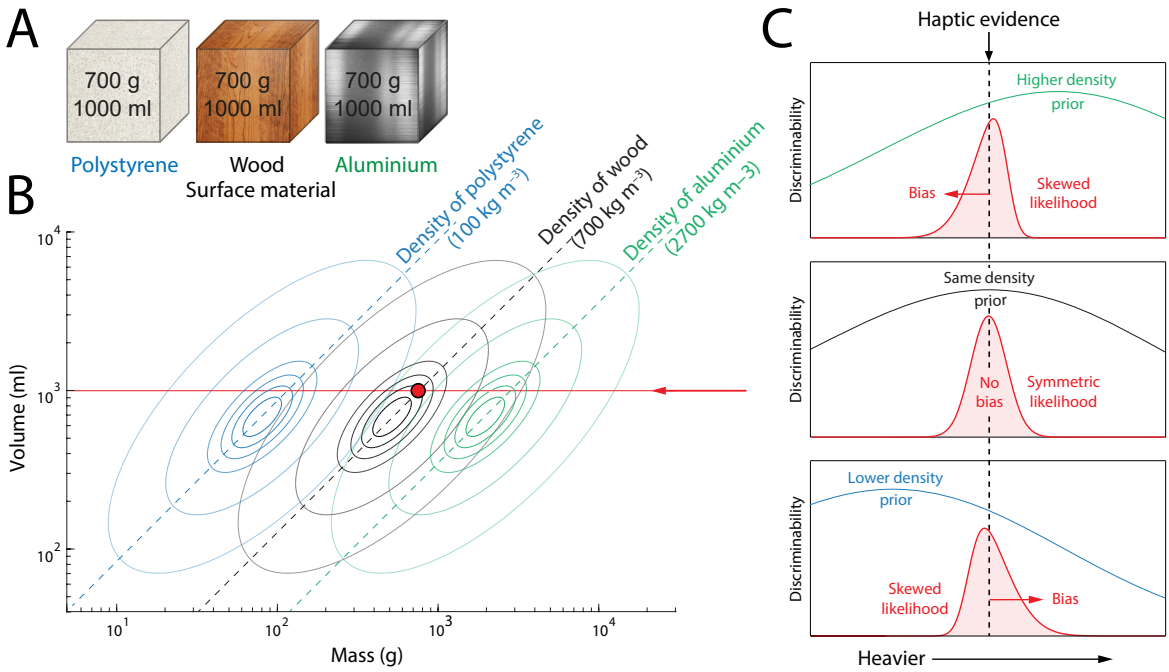


Figure 4: Predicted effects of surface material on mass estimation. (A) Stimuli consisting of three cubes of identical size and weight but different surface material. (B) Contours illustrate possible prior distributions over mass and volume based on past experience of objects made from the three different materials. In the absence of haptic feedback, the expected mass-volume ratio of the three objects will approximately match the density of the surface material (dashed lines). The red disc corresponds to the actual mass and volume of all three objects. (C) Illustrates estimation of mass for a lifted object with an expected density that is larger (top), equal (middle) or lower (bottom) than its actual density. Discriminability varies in proportion to the conditional probability of object mass, given visual evidence of object volume and object density. For the same haptic feedback (black dashed line) the different gradients of discriminability lead to different likelihood functions (red curves) and different biases (red arrows) in the mean posterior estimates.

Material-weight illusion

We next considered whether the efficient coding principle could explain the material-weight illusion (MWI) (Harshfield & DeHardt, 1970), in which an object with the visual appearance of a dense material is perceived on lifting as lighter than a matched object of seemingly less dense material. We assumed that the visual appearance of different surface materials (Fig 4A; objects used in Buckingham et al., 2009) would invoke different prior distributions based on previous experience of objects made of that material, as illustrated in Fig 4B. Specifically, we assumed the induced prior would reflect a mean density matching that of the surface material (dashed lines indicate combinations of mass and volume consistent with the density of each material).

As illustrated in Fig 4C, the consequence of efficiently encoding object mass conditioned on its volume and surface material depends on the difference between density of the surface material and density of the object. If the surface material is of higher density than the object as a whole (top panel), the prior distribution conditioned on volume will peak at a higher mass than the one indicated by haptic evidence from lifting the object (vertical dashed line), resulting in a gradient of discriminability (green line) that biases the mean posterior estimate towards lower masses (red arrow). The converse holds for a surface material of lower density than the

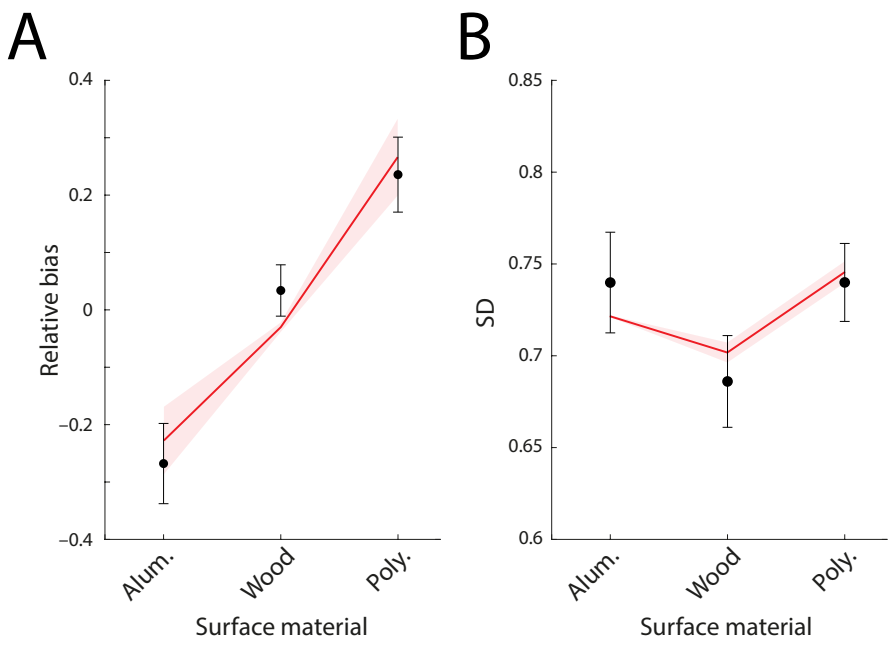


Figure 5: Material-weight illusion data and fits. Black circles indicate mean relative bias (A) and mean within-participant SD (B) of human observers in estimating mass of lifted objects with different surface materials but identical mass and volume (Buckingham et al., 2009). Error bars indicate ±1 within-subject SEM. Red lines show predictions of the fitted efficient coding model.

object (bottom panel). For a surface material that matches the object density (middle panel), the prior probability, and hence the discriminability, peaks at the true mass of the object, so no bias is induced. Additionally, the discriminability of mass in this case is predicted to be greater than when the surface material and object density are in conflict.

Fig. 5A plots relative bias in magnitude judgments of weight when lifting the objects illustrated in Fig. 4A. Consistent with the MWI, the data show a positive bias in estimated mass for less compared to more dense surface materials, despite all the objects having the same mass and volume. Fig. 5B shows the corresponding within-participant SDs (maximum likelihood parameters, median values: prior variance, $s^2 = 149$; bias constant, $c_b = 0.0093$; s.d. constant, $c_\sigma = 0.0217$).

Discussion

We have shown how weight illusions arise as an indirect consequence of efficient coding based on everyday experience with objects, in which variation in the discriminability of object properties, reflecting their relative prevalence in the environment, induces estimation biases when haptic evidence for a lifted object's mass is uncertain. This account of the SWI as an adaptation of encoding to environmental statistics may also explain why prolonged exposure to objects for which size and weight were anti-correlated caused a reversal of the illusion (Flanagan et al., 2008). This would match the pattern of adaptation recently demonstrated for orientation biases after exposure to stimuli drawn from a distribution that favoured oblique angles (Taylor & Bays, 2018), and implies that efficient coding is not an immutable feature of sensory processing but based on continual learning of environmental statistics.

This is not the first study to propose that the seemingly "anti-Bayesian" bias of the SWI might be consistent with Bayesian estimation. Peters et al. (2016) suggested that, when com-

paring two objects A and B, the brain evaluates three categorical hypotheses about their relative densities: that they have equal density, that A is more dense than B, or that A is less dense than B. Each hypothesis is associated with a different prior distribution over relative volume and density, and the influence of these competing priors leads to a bias in the posterior estimate of relative weight in the direction of the SWI. However, the categorical treatment of relative density is not clearly justified in this account and appears to be largely *ad hoc*. Moreover, while the categorical model with hand-tuned parameters reproduced the magnitude and overall pattern of decreasing SWI with decreasing volume ratio, it failed to predict the empirical differences in SWI for object pairs with the same volume ratio (marked by equals signs in Fig 3A) nor the effects of object mass on SWI (compare panels left to right in Fig 3A). The categorical model also makes no explicit predictions for variability of estimates, and its quantitative predictions were not evaluated or compared to data. The present model reproduced all of these elements of the SWI based on the ecological goal of efficient coding and a single prior based on natural statistics.

The present results do not uniquely specify a coding scheme for the object properties, and a number of different encodings compatible with the efficient coding goal are at least qualitatively consistent with the observed biases and variability in estimation (Fig. S1). A simple scheme would be to encode an object's log-weight in relative terms, as a deviation from its expected log-weight given object volume. Allocating coding resources (e.g. neuronal tuning density) preferentially to smaller deviations would achieve efficiency with respect to natural statistics (note this scheme depends on properties of the bivariate Gaussian prior and might not be realizable for a different prior distribution). Early accounts of weight illusions frequently appealed to the contrast between expected (based on visual properties) and actual sensory feedback experienced during the lift as the basis of the illusion (Davis & Roberts, 1976; Ross, 1969), i.e., a larger object feels lighter when lifted because it is lighter than *expected*. While this informal account is only descriptive of the illusion, the present results present a normative basis for biasing perception relative to the expected mass.

The SWI (Buckingham & Goodale, 2010) and MWI (Buckingham et al., 2011) can be induced by preceding visual observation of an object's size even if visual and haptic evidence of size are prevented during the lift. The magnitude of both illusions was found to be weaker in this scenario than with concurrent vision, consistent with greater uncertainty about object volume (SWI) or density (MWI) leading to a broadening of the prior distribution with respect to object mass. According to the efficient coding model, bias magnitude is proportional to the gradient (rate of change) of prior probability with respect to mass: when the prior is broader and flatter (increased s.d.) the strength of bias is expected to decrease. Similarly, Ellis and Lederman (1993) observed a weaker SWI based on visual evidence of size alone than with both visual and haptic feedback, consistent with the efficient coding account. They also observed that the SWI based on haptics alone was stronger than with vision only, suggesting that haptic feedback obtained from holding an object supported a more precise estimate of the object's volume than visual observation (Gepshtein & Banks, 2003; Hershberger & Misceo, 1996).

The SWI persists over repeated lifting of the same objects, but anticipatory motor responses, in the form of grip and load forces, adapt over only a few lifts to become appropriately scaled for veridical object weights (Brayanov & Smith, 2010; Flanagan & Beltzner, 2000). This dissociation may reflect the different goals of perceptual and sensorimotor systems. Whereas the SWI can be interpreted as the perceptual system attempting to minimize estimation error while taking into account anisotropy in encoding fidelity, the sensorimotor system's goal is to apply the correct forces to smoothly lift each object without slipping. Lift kinematics and haptic feedback provide strong error signals about inappropriate grasp that can be used for corrective adjustment of force on the next lift of the same object, and iterated until the desired kinematics are achieved. In contrast, the SWI is a bias in relative judgements of weight that, according to the present account, reflects optimal perception given the combination of haptic and visual

feedback experienced during a lift. As such, there is no error signal generated by repeated lifting that would lead this perception to change.

Methods

Estimating natural statistics

We estimated the joint distribution of experienced object mass and volume by collating datasets 1–4 of Peters et al. (2015), consisting of measured properties of a sampling of everyday objects. We excluded unliftable objects (dataset 5) on the basis that the mass of such objects would not be expected to contribute to prior expectation for lifted objects, the basis of the SWI.

Efficient coding model of the SWI

We assumed a bivariate normal joint prior distribution over log-mass (m) and log-volume (v)

$$\begin{bmatrix} m \\ v \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} \mu_m \\ \mu_v \end{bmatrix}, \begin{bmatrix} \sigma_m^2 & \sigma_{mv} \\ \sigma_{mv} & \sigma_v^2 \end{bmatrix} \right). \tag{1}$$

The prior distribution of log-mass conditioned on log-volume is then given by

$$m|v \sim \mathcal{N} \left(\mu_m + \frac{\sigma_{mv}}{\sigma_v^2}(v - \mu_v), \sigma_m^2 - \frac{\sigma_{mv}^2}{\sigma_v^2} \right). \tag{2}$$

Following the formulation of Wei and Stocker (2017), an efficient coding of log-mass of an object with respect to its log-volume will result in a discrimination threshold or just-noticeable difference for log-mass,

$$D_m(m, v) \propto \frac{1}{p(m|v)}, \tag{3}$$

and a Bayesian estimate of log-mass based on this encoding will have bias,

$$b_m(m, v) \propto \left(D_m(m, v)^2 \right)' \tag{4}$$

with this and all derivatives taken with respect to m. The sign of the proportionality is positive (i.e. repulsive) for the posterior mean estimate (minimizing squared error) and all estimates that minimize loss functions with exponent ≥ 1. The s.d. of the Bayesian estimate is

$$\hat{\sigma}_m(m, v) \propto D_m(m, v)(1 + b'_m(m, v)). \tag{5}$$

We fit this model to data from the SWI study of Peters et al., 2015. Stimuli were twelve cubes comprising three different masses and four different volumes. Participants lifted objects in pairs of the same mass but different volumes, and reported an estimate of the ratio of their masses. We assumed that the estimated log-mass $\hat{m}u_i$ of an object with log-mass m_i and log-volume v_j is normally distributed with bias b_{ij} and s.d. $\hat{\sigma}_{ij}$ given by Eqs. 4 & 5 above,

$$\widehat{m}_{ij} \sim \mathcal{N}(m_i + b_{ij}, \hat{\sigma}_{ij}^2), \tag{6}$$

in which case the log of the estimated mass ratio of two objects with the same log-mass m_i and log-volumes v_j and v_k is then also normal with

$$\log \hat{r}_{ijk} = \widehat{m}_{ij} - \widehat{m}_{ik} \sim \mathcal{N}(b_{ij} - b_{ik}, \hat{\sigma}_{ij}^2 + \hat{\sigma}_{ik}^2). \tag{7}$$

Implementing the model

The model as described above is overparameterized (i.e. different combinations of parameters make identical predictions), so we re-write the prior density conditional on volume (Eq. 2) as

$$p(m|v) = \phi(m; \beta v + c_m, s) \tag{8}$$

with $\phi(x; \mu, \sigma)$ the normal p.d.f. evaluated at x. The new parameters with respect to those of Eq. 2 are

$$\beta = \frac{\sigma_{mv}}{\sigma_v^2}, \tag{9}$$

$$c_m = \mu_m - \frac{\sigma_{mv}}{\sigma_v^2}\mu_v = \mu_m - \beta\mu_v, \tag{10}$$

$$s^2 = \sigma_m^2 - \frac{\sigma_{mv}^2}{\sigma_v^2}. \tag{11}$$

Defining c_b and c_σ as proportionality constants in Eqs. 4 & 5, we have bias

$$b_m(m, v) = c_b(p(m|v)^{-2})' \tag{12}$$

and s.d.

$$\begin{aligned} \hat{\sigma}_m(m, v) &= c_\sigma(1 + b'_m(m, v))p(m|v)^{-1} \\ &= c_\sigma(1 + c_b(p(m|v)^{-2})'')p(m|v)^{-1} \end{aligned} \tag{13}$$

We used non-linear optimization (direct search) to find maximum likelihood values for each participant of parameters β, c_m, and s^2, the slope, intercept and variance describing the prior distribution, and c_b and c_σ, the constants of proportionality for bias and s.d. respectively. Parameters of the illustrative bivariate normal prior shown in Fig. 2 (contours) were medians calculated from maximum likelihood fitted parameter values with $\mu_v = 6, \sigma_{mv} = 5$.

Efficient coding model of the MWI

We extend Eq. 2 in our model of the SWI to describe the prior distribution of log-mass conditioned on both log-volume and surface material as

$$m|v, M \sim \mathcal{N}\left(v + \bar{d}(M), s^2\right) \tag{14}$$

where $\bar{d}(M)$ is the mean log-density of material M.

We fit data from the study of Buckingham et al. (2009), which compared the perceived weights of lifted objects that appeared to be made of different materials. Stimuli consisted of three cubes, each of a different surface material: aluminium (of density 2.7 g cm^{-3}), wood (0.7 g cm^{-3}) and expanded polystyrene (0.1 g cm^{-3}). All three cubes were identical in mass (700 g) and volume (1,000 cm^3), having the density of wood (0.7 g cm^{-3}). Participants lifted each cube 15 times in a randomized sequence of triplets, reporting an absolute magnitude estimate of object weight after each lift. Each participant's magnitude estimates were z-scored in the original study (scaled and shifted to have pooled mean of zero and s.d. of 1). The authors noted a consistent tendency for magnitude estimates to increase over the course of the experiment, so we additionally detrended the data by subtracting the mean rating from each triplet of lifts in the sequence, before calculating the s.d. of estimates for each object.

For modelling purposes, we assume that the magnitude estimates, ψ, are linearly related to the perceived log-mass on each lift, i.e. $\psi_i = \gamma\widehat{m}_i + c$, with $\widehat{m}_i \sim \mathcal{N}(m + b_i, \hat{\sigma}_i^2)$ as above (Eq. 6), and bias b_i and variability $\hat{\sigma}_i$ for each material obtained from the corresponding prior (Eq. 14) as in Eqs. 4 & 5. We used non-linear optimization (direct search) to find maximum likelihood parameters s^2, the variance of the conditional prior density; c_b and c_σ, constants of proportionality for bias and s.d. respectively.

Acknowledgments

We thank Máté Lengyel for helpful discussion, and Megan Peters and Gavin Buckingham for sharing data. This work was funded by the Wellcome Trust (grant 106926).

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, *61*(3), 183. https://doi.org/10.1037/h0054663

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, *1*(01).

Brayanov, J. B., & Smith, M. A. (2010). Bayesian and "anti-Bayesian" biases in sensory integration for action and perception in the size–weight illusion. *Journal of neurophysiology*, *103*(3), 1518–1531. https://doi.org/10.1152/jn.00814.2009

Buckingham, G., Cant, J. S., & Goodale, M. A. (2009). Living in A Material World: How Visual Cues to Material Properties Affect the Way That We Lift Objects and Perceive Their Weight. *Journal of Neurophysiology*, *102*(6), 3111–3118. https://doi.org/10.1152/jn.00515.2009

Buckingham, G., & Goodale, M. A. (2010). Lifting without Seeing: The Role of Vision in Perceiving and Acting upon the Size Weight Illusion. *PLOS ONE*, *5*(3), e9709. https://doi.org/10.1371/journal.pone.0009709

Buckingham, G., Ranger, N. S., & Goodale, M. A. (2011). The material–weight illusion induced by expectations alone. *Attention, Perception, & Psychophysics*, *73*(1), 36–41. https://doi.org/10.3758/s13414-010-0007-4

Davis, C. M., & Roberts, W. (1976). Lifting movements in the size-weight illusion. *Perception & Psychophysics*, *20*(1), 33–36. https://doi.org/10.3758/BF03198701

Ellis, R. R., & Lederman, S. J. (1993). The role of haptic versus visual volume cues in the size-weight illusion. *Perception & Psychophysics*, *53*(3), 315–324. https://doi.org/10.3758/BF03205186

Flanagan, J. R., & Beltzner, M. A. (2000). Independence of perceptual and sensorimotor predictions in the size–weight illusion. *Nature Neuroscience*, *3*(7), 737–741. https://doi.org/10.1038/76701

Flanagan, J. R., Bittner, J. P., & Johansson, R. S. (2008). Experience Can Change Distinct Size-Weight Priors Engaged in Lifting Objects and Judging their Weights. *Current Biology*, *18*(22), 1742–1747. https://doi.org/10.1016/j.cub.2008.09.042

Ganguli, D., & Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural computation*. https://doi.org/10.1162/NECO_a_00638

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*(6), 483–488.

Harshfield, S. P., & DeHardt, D. C. (1970). Weight judgment as a function of apparent density of objects. *Psychonomic Science*, *20*(6), 365–366. https://doi.org/10.3758/BF03335692

Hershberger, W. A., & Misceo, G. F. (1996). Touch dominates haptic estimates of discordant visual-haptic size. *Perception & Psychophysics*, *58*(7), 1124–1132. https://doi.org/10.3758/BF03206838

Morais, M., & Pillow, J. W. (2018). Power-law efficient neural codes provide general link between perceptual bias and discriminability. *NeurIPS*, 5076–5085.

Murray, D. J., Ellis, R. R., Bandomir, C. A., & Ross, H. E. (1999). Charpentier (1891) on the size—weight illusion. *Perception & Psychophysics*, *61*(8), 1681–1685. https://doi.org/10.3758/BF03213127

Peters, M. A. K., Balzer, J., & Shams, L. (2015). Smaller = Denser, and the Brain Knows It: Natural Statistics of Object Density Shape Weight Expectations. *PLOS ONE*, *10*(3), e0119794. https://doi.org/10.1371/journal.pone.0119794

Peters, M. A. K., Ma, W. J., & Shams, L. (2016). The Size-Weight Illusion is not anti-Bayesian after all: A unifying Bayesian account. *PeerJ*, *4*, e2124. https://doi.org/10.7717/peerj.2124

Ross, H. E. (1969). When is a weight not illusory? *Quarterly Journal of Experimental Psychology*, *21*(4), 346–355. https://doi.org/10.1080/14640746908400230

Taylor, R., & Bays, P. M. (2018). Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *Journal of Neuroscience*, 1018–18. https://doi.org/10.1523/JNEUROSCI.1018-18.2018

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Wei, X.-X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, 201619153. https://doi.org/10.1073/pnas.1619153114

Yerxa, T. E., Kee, E., DeWeese, M. R., & Cooper, E. A. (2020). Efficient sensory coding of multidimensional stimuli. *PLOS Computational Biology*, *16*(9), e1008146. https://doi.org/10.1371/journal.pcbi.1008146

Zhu, Q., & Bingham, G. P. (2011). Human readiness to throw: The size–weight illusion is not an illusion when picking the best objects to throw. *Evolution and Human Behavior*, *32*(4), 288–293. https://doi.org/10.1016/j.evolhumbehav.2010.11.005

Supplementary Information

Multivariate efficient coding model

In the univariate case considered by Wei and Stocker (2015, 2017), the efficient coding goal of maximizing mutual information is attained by setting Fisher Information, $J(\theta)$, to a function of the prior,

$$J(\theta) \propto p(\theta)^2. \tag{15}$$

The methods in the main text extend the efficient coding model to the case of correlated variables (mass and volume) by considering the conditional probability of mass with respect to volume, implicitly assuming that sensory evidence for volume contributes to the efficient representation of mass but is encoded independently. Here we consider how the efficient coding principle could be extended to joint encoding of the two variables with respect to their prior distribution.

The equivalent constraint to Eq. 15 for the multivariate case is

$$\det \mathbf{J}(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta})^2, \tag{16}$$

but note this does not fully constrain the Fisher Information matrix, with the implication that there is more than one encoding of a combination of variables that satisfies the efficient coding constraint. Here we present one formulation of the model that satisfies this equation, based on factorizing the prior distribution into independent components.

We make a change of basis,

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{S}^{-1}\mathbf{R}^{-1}\left(\begin{bmatrix} m \\ v \end{bmatrix} - \begin{bmatrix} \mu_m \\ \mu_v \end{bmatrix}\right) \tag{17}$$

based on the eigendecomposition of the prior's covariance matrix,

$$\boldsymbol{\Sigma} = \mathbf{RSSR}^{-1} \tag{18}$$

such that prior probabilities $p(x) = \phi(x)$ and $p(y) = \phi(y)$ are independent standard normal distributions, with $p(x, y) = p(x)p(y)$, and prior densities in the two bases related through

$$p_{M,V}(m, v) = p_{X,Y}(x, y) \det \mathbf{D} \tag{19}$$

where $\mathbf{D} = \frac{\partial(x,y)}{\partial(m,v)}$ is the Jacobian of the transformation which in this case is simply $\mathbf{D} = \mathbf{RS}$. Setting Fisher Information for each independent variable according to the univariate constraint (Eq. 15) leads to the joint Fisher Information matrix,

$$\mathbf{J}_{XY} \propto \begin{bmatrix} p(x)^2 & 0 \\ 0 & p(y)^2 \end{bmatrix} \tag{20}$$

$$\implies \det \mathbf{J}_{XY} \propto p(x)^2 p(y)^2 = p(x, y)^2. \tag{21}$$

The FI matrix for the original variables is

$$\mathbf{J}_{MV} = \mathbf{D}^{\mathrm{T}}\mathbf{J}_{XY}\mathbf{D} \tag{22}$$

$$\implies \det \mathbf{J}_{MV} = (\det \mathbf{D})^2 \det \mathbf{J}_{XY} \tag{23}$$

$$\propto (\det \mathbf{D})^2 p(x, y)^2 \tag{24}$$

$$\implies \det \mathbf{J}_{MV} \propto p(m, v)^2 \tag{25}$$

satisfying Eq. 16.

Prior probabilities of x and y are mutually independent, so we can base predictions for efficient encoding of x (or equivalently y) on the univariate case, i.e.,

$$b_x(x) \propto \frac{d}{dx}\frac{1}{p(x)^2} = -2\frac{p'(x)}{p(x)^3} = 2\frac{x}{\phi(x)^2} \tag{26}$$

$$\Rightarrow b'_x(x) \propto \frac{2+4x^2}{\phi(x)^2} \tag{27}$$

and s.d.

$$\hat{\sigma}_x(x) \propto \frac{1+b'_x(x)}{p(x)} = \frac{1}{\phi(x)} + \frac{2+4x^2}{\phi(x)^3} \tag{28}$$

Bias in the original variables can be obtained as

$$\begin{bmatrix} b_m(m) \\ b_v(v) \end{bmatrix} = \mathbf{RS} \begin{bmatrix} b_x(x) \\ b_y(y) \end{bmatrix} \tag{29}$$

and s.d. from

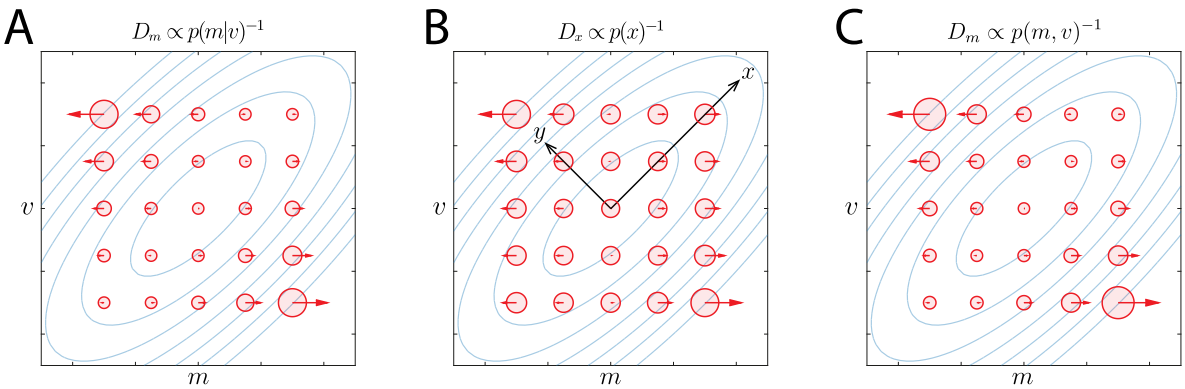
$$\mathrm{Cov}(m,v) = \begin{bmatrix} \hat{\sigma}^2_m(m) & \hat{\sigma}(m,v) \\ \hat{\sigma}(m,v) & \hat{\sigma}^2(v) \end{bmatrix} = \mathbf{RS} \begin{bmatrix} \hat{\sigma}^2_x(x) & 0 \\ 0 & \hat{\sigma}^2_y(y) \end{bmatrix} \mathbf{SR}^{\mathrm{T}} \tag{30}$$

Predictions for this model are illustrated in Fig. S1B. Predictions for the main model are duplicated in Fig. S1A for comparison.

Joint density model

The model in the main text allocates resources for encoding object mass according to the conditional prior probability of mass given object volume, i.e. $J_M \propto p(m|v)$. This is an appropriate solution if the total resources available to encode mass are constant irrespective of volume, e.g., the case of distributing preferred values in a neural population code. This also allows for the simple coding scheme discussed in the main text, in which a population with fixed tuning encodes deviations from the expected mass given object volume.

An alternative is that the total resource for encoding mass can update according to volume, for example, gain modulation in a population code where the goal is to maximize mutual information under an energetic constraint. In such cases, an appropriate solution would be to allocate resources according to the joint prior probability of mass and volume, $J_M \propto p(m,v)$, thereby conserving activity for those objects with the most frequently encountered combinations of properties. Predictions for this model are illustrated in Fig. S1C.



Supplementary Figure 1: Predictions of alternative encoding models. Predicted SD (circle diameters) and bias (arrows) in estimates of log-mass for objects with a range of true log-volumes (v) and log-masses (m), for (A) the model in the main text, where object mass is efficiently encoded with respect to its conditional prior probability given object volume; (B) a model where object mass and volume are jointly efficiently encoded with respect to the eigenvectors (x and y) of the prior distribution; (C) a model where object mass is efficiently encoded with respect to the joint prior probability of mass and volume.