Contribution of surface material and size to the expected versus the perceived weight of objects

Michele Vicovaro¹ · Luigi Burigana¹

Published online: 14 October 2016
© The Psychonomic Society, Inc. 2016

Abstract Because the perceived weight of objects may be affected by various nonweight properties, such as their size and the density of their surface material, relative weight is sometimes misperceived (the *size–weight illusion* and the *material–weight illusion*, respectively). A widely accepted explanation for weight illusions is provided by the so-called *expectation model*, according to which the perceived weight stems from the contrast between the actual and expected weights. In the present study, we varied both the surface material and the size of stimuli, while keeping constant their physical weights. In Experiment 1, the participants lifted the stimuli by grasping them on opposite sides, whereas in Experiment 2 they lifted them by using a string that was attached to their top surface. We used a variant of the *random conjoint measurement* paradigm to obtain subjective interval scales of the contributions of surface material and size to the expected and the perceived weight of the stimuli. Inconsistently with the predictions from the expectation model, we found, in both experiments, that the surface material contributed more than the size to the expected weight, whereas the size contributed more than the surface material to the perceived weight. The results support the hypothesis that perceived weight may depend on implicit, rather than explicit, weight expectations.

Keywords Weight perception · Size–weight illusion · Material–weight illusion · Information integration · Conjoint measurement · Haptics · Multisensory processing · Model selection

The *size–weight illusion* (SWI) refers to the phenomenon whereby, when two objects of the same physical weight but different sizes are lifted, the smaller object usually feels heavier than the larger object (Charpentier, 1891; see also Nicolas, Ross, & Murray, 2012). The *material–weight illusion* (MWI) refers instead to the phenomenon whereby, when two objects of the same physical weight but different surface materials are lifted, the object with the less dense surface material (e.g., polystyrene) usually feels heavier than the object with the denser surface material (e.g., iron; Harshfield & DeHardt, 1970; Seashore, 1899).

Perhaps the most widely accepted explanation for the SWI and MWI is provided by the so-called *expectation model* (Flanagan, Bittner, & Johansson, 2008; Ross, 1969), according to which the SWI and the MWI subtend top-down perceptual processes. Over their lifetime, people learn that larger objects are typically heavier than smaller objects, and that objects made of denser materials are typically heavier than objects made of less dense materials. Therefore, people expect larger objects to weigh more than smaller objects, and they expect objects with denser surface materials to weigh more than objects with less dense surface materials. Perceived weight thus results from the contrast between the sensory information about the actual weight of an object and the cognitive information about its expected weight.¹ Because of that contrast, if two objects of the same physical weight are expected to have different weights, then the object that is expected to be heavier will be perceived to be lighter. Support for the expectation

---

✉ Michele Vicovaro
michele.vicovaro@unipd.it

¹ Department of General Psychology, University of Padua, via Venezia 8, I-35131 Padova, Italy

¹ After an empirical study, Plaisier and Smeets (2012) suggested that the SWI should instead be called the “size–mass illusion.” However, here we refer to the “size–weight illusion” for the sake of consistency with other weight illusions, such as the MWI and the golf-ball illusion.

model has been provided by studies showing that weight expectations generated prior to the lifting action can directly cause the SWI (Buckingham & Goodale, 2010; cf. Masin & Crestoni, 1988) and the MWI (Buckingham, Ranger, & Goodale, 2011), as well as by studies showing that the magnitude of the SWI is inversely related to the amount of participants' experience with objects that reverse the natural size–weight correlation (Flanagan et al., 2008; Nakatani, 1985).

The expectation model provides a unified explanation for various weight illusions (see Buckingham, 2014, for a review), such as the SWI, the MWI, and the so-called *golf-ball illusion*. Ellis and Lederman (1998) asked golfers and nongolfers to judge the expected weight of regular and practice golf balls. Golfers correctly expected the regular balls to be heavier than the practice balls, whereas nongolfers, who lacked experience with the two kinds of balls, believed that they had approximately the same weight. The participants were then asked to judge the perceived weight of regular and practice golf balls that were artificially equated in physical weight. The results showed that golfers, who expected the regular balls to be heavier, perceived them to be lighter than the practice balls (i.e., the golf-ball illusion), whereas nongolfers, who expected the two kinds of balls to weigh the same, perceived them to weigh the same. The golf-ball illusion constitutes an outstanding example of the top-down effects of prior weight expectations on the perceived weights of objects.

Early researchers attempted to ground the contrast between actual and expected weight on relatively simple physiological mechanisms. Specifically, some students (Davis & Roberts, 1976; Gordon, Forssberg, Johansson, & Westling, 1991) hypothesized that objects that are lifted more easily would be perceived as lighter than objects that are lifted with more difficulty. Because objects that are expected to be heavier are lifted with greater force than objects of the same physical weight that are expected to be lighter, the former would be lifted more easily, and thus, they would be perceived to be lighter than the latter. Unfortunately, this relatively straightforward explanation of the effects of prior weight expectations on weight perception is probably incorrect. It has indeed been found that the lifting force rapidly adapts to the actual weight of objects; that is, it ceases to be affected by prior weight expectations after a few lifting trials. Despite this, weight illusions persist even after several lifts of the same set of objects (Brayanov & Smith, 2010; Flanagan & Beltzner, 2000; Flanagan et al., 2008; Grandy & Westwood, 2006). This finding suggests that the contrast between actual and expected weight is probably due to central perceptual–cognitive processes, rather than to peripheral physiological mechanisms.

From manipulation of one to manipulation of two nonweight stimulus properties: implications for the expectation model

Studies on illusory weight perception have typically been conducted by varying the physical weight of stimuli and one of their nonweight properties (e.g., size or surface material). However, researchers have recently started exploring the relationship between expected and perceived weight in studies in which two or more nonweight properties of the stimuli are jointly manipulated (Buckingham & Goodale, 2013; Buckingham & MacDonald, 2016; Dijkers, 2008). Let us denote by w the actual weight of objects, by ε their expected weight, and by π their perceived weight. Furthermore, let us suppose that x and y are two nonweight properties of the stimuli (i.e., properties different from w) that may be conveniently manipulated in experiments. Let $\xi_\varepsilon(x)$ and $v_\varepsilon(y)$ be real-valued functions (to be estimated) quantifying how the properties x and y influence the expected weight ε of the stimuli. Similarly, let $\xi_\pi(x)$ and $v_\pi(y)$ be two real-valued functions quantifying how the same properties influence the perceived weight π of the stimuli. We presume the following additive combination functions: $\varepsilon(x, y) = \xi_\varepsilon(x) + v_\varepsilon(y)$ and $\pi(w, x, y) = \omega(w) + \xi_\pi(x) + v_\pi(y)$, where $\omega(w)$ is a real-valued function quantifying the contribution of w to perceived weight. Consistent with the basic features of the expectation model, we also presume $\pi(w, x, y) = \omega(w) - \varepsilon(x, y) = \omega(w) - \xi_\varepsilon(x) - v_\varepsilon(y)$; that is, perceived weight depends on the contrast between the sensory information about actual weight and expected weight. The two equations above for $\pi(w, x, y)$ imply the following equation:

$$\xi_\pi(x) + v_\pi(y) = -\xi_\varepsilon(x) - v_\varepsilon(y). \quad (1)$$

In real experimental contexts, testing whether Eq. 1 holds true may constitute a valuable test of the validity of the expectation model. In this regard, we highlight two important considerations. The first is that the pair of functions $\xi_\pi(x)$ and $v_\pi(y)$ (left-hand side of Eq. 1) and the pair of functions $\xi_\varepsilon(x)$ and $v_\varepsilon(y)$ (right-hand side of Eq. 1) refer to different psychological variables (i.e., perceived and expected weight, respectively), and thus they should be estimated from different data sets. The second consideration is that the estimates of these functions may reasonably be defined, at best, as scale values on interval measurement scales—that is, scale values well determined “up to linear transformations.” Taken together, the two considerations imply that the measurement unit and the origin of estimated scale values $\xi_\pi(x)$ and $v_\pi(y)$ may be somewhat arbitrary, and may be different from those of the estimated scale values $\xi_\varepsilon(x)$ and $v_\varepsilon(y)$. For the sake of argument, we note that testing whether $\xi_\pi(x) = -\xi_\varepsilon(x)$ and $v_\pi(y) = -v_\varepsilon(y)$ would not constitute a reliable test of Eq. 1 because the outcome of this test would also depend on arbitrary features of the scale values involved. In the following paragraphs we will

discuss three possible tests of Eq. 1 that are independent of such arbitrary features. We will discuss these tests in terms of distinct *predictions*, that should be met if Eq. 1—and thus the expectation model—are correct.

Prediction 1 For any fixed w , $\xi_\pi(x)$ and $v_\pi(y)$ should be decreasing monotone functions if and only if $\xi_\varepsilon(x)$ and $v_\varepsilon(y)$ are increasing monotone functions. In other words, the perceived weight of stimuli should decrease with x and y if and only if the expected weight of the stimuli increases with x and y .

Prediction 2 For any fixed value of w , any pair (x_1, x_2) of values of x , and any pair (y_1, y_2) of values of y , the relative contribution of x and y to perceived weight should be the same (within the two pairs of values) as their relative contribution to perceived weight (i.e., $[\xi_\pi(x_1) - \xi_\pi(x_2)]/[v_\pi(y_1) - v_\pi(y_2)] = [\xi_\varepsilon(x_1) - \xi_\varepsilon(x_2)]/[v_\varepsilon(y_1) - v_\varepsilon(y_2)]$).

Prediction 3 In a set of stimuli that vary in x and y , if two stimuli have the same physical weight, then the stimulus that is expected to be heavier should be perceived to be lighter than the other stimulus (i.e., for any two stimuli A and B , if $w(A) = w(B)$ and $\varepsilon(A) > \varepsilon(B)$, then $\pi(A) < \pi(B)$).

To our knowledge, Predictions 1–3 have never been formalized and tested in a single study, though they have been separately tested in different studies. Consistently with Prediction 1, Dijker (2008) showed that whereas the perceived weight of toy dolls decreased with their apparent size and with their apparent physical strength, their expected weight increased along with both factors. Similarly, Buckingham and Goodale (2013) found that whereas the perceived weight of cubic boxes decreased with their size and with the density of their surface material, their expected weight increased along with both factors. To our knowledge, a clear violation of Prediction 1 has never been reported in the literature on illusory weight perception. Prediction 2 has been previously tested and disconfirmed in Dijker's (2008) study, which showed that whereas the expected weight of toy dolls was affected more by their apparent size than by their apparent physical strength, their perceived weight was affected more by their apparent physical strength than by their apparent size. Prediction 3 has typically been confirmed in studies in which only one nonweight property of the stimuli was manipulated, such as the density of the surface material (Harshfield & DeHardt, 1970) or the brightness (Walker, Francis, & Walker, 2010). However, Buckingham and MacDonald (2016) presented the participants in their study with three objects that had approximately the same physical weight, but that differed in various nonweight properties (i.e., a golf ball, a foam soccer ball, and an inflated beach ball). The results showed that, inconsistently with Prediction 3, the golf ball was both expected and perceived to be heavier than the foam

soccer ball, and the latter was both expected and perceived to be heavier than the inflated beach ball.

Apart from Predictions 1–3, the expectation model has also been tested through the comparison of the algebraic form of functions describing how size and surface material combine in affecting expected and perceived weight of objects. Buckingham and Goodale (2013) presented the participants in their study with four cubic boxes of the same physical weight, varying in surface material and size according to a 2×2 factorial design (cf. Ross, 1969). They found that the expected weight of the stimuli increased with the product of their size and the density of their surface material, whereas their perceived weight decreased with the sum of the two experimental factors. This finding was inconsistent with the predictions from the expectation model, according to which, if expected weight increases with the product of two experimental factors, then perceived weight should decrease with the product, rather than with the sum, of the two experimental factors.

In sum, the results of empirical studies appear to suggest that, when two or more nonweight properties of stimuli are jointly manipulated in the same experiment, some predictions from the expectation model may fail to be met. This finding has cast doubts on the idea that the perceived weight of objects may simply be explained in terms of the contrast between the actual and the expected weight. Various alternative hypotheses have been proposed (see Buckingham, 2014; Buckingham & Goodale, 2013; Buckingham & MacDonald, 2016; Dijker, 2008, 2014), to which we will return below. We wish to emphasize here that, when two nonweight properties of the stimuli are jointly manipulated in a single experiment, testing Predictions 1–3 appears to constitute a promising strategy for defining the extent and the limits of the expectation model.

In the present study, we varied the surface material and the size of the stimuli (i.e., cubic boxes) according to a 3 (surface material) $\times 5$ (size) factorial design, while keeping constant their physical weight. We used a variant of the *random conjoint measurement* paradigm (RCM; Falmagne, 1976; Ho, Landy, & Maloney, 2008; Knoblauch & Maloney, 2012, chap. 8) in order to obtain subjective interval scales of the contribution of surface material and size to the expected and the perceived weight of the stimuli, which allowed us to test Predictions 1–3 from the expectation model. Our contribution bears some resemblance to Buckingham and Goodale's (2013) study, though here we tested crucial predictions of the expectation model that are independent of the correct assessment of the additive versus multiplicative forms of the combination functions.²

² For instance, the three predictions would also hold true if we assume $\varepsilon(x, y) = \xi_\varepsilon(x) \times v_\varepsilon(y)$, $\pi(w, x, y) = \xi_\pi(x) \times v_\pi(y)$, and $\pi(w, x, y) = \omega(w)/\varepsilon(x, y)$.

Experiment 1

Method

Participants Ten participants (six females, four males) took part in the experiment. They were undergraduate or graduate students at the Faculty of Psychology, University of Padua, ranging in age from 20 to 30 years ($M = 24.4$, $SD = 4.17$). All had normal or corrected-to-normal vision and normal function of their upper limbs. They were naive as to the purposes of the experiment and received €20 for their participation.

Stimuli The stimuli were 15 cubic boxes each weighing 1.2 kg, which were built according to a 3 (Surface Material) $\times$ 5 (Size) factorial design. The surface material of the stimuli could be polystyrene (density = 0.017 g/cm^3), wood (0.44 g/cm^3), or black-painted solid clay (1.44 g/cm^3). The sides of the stimuli varied from 14 to 26 cm in five uniform, 3-cm steps (see Fig. 1). The boxes were made of an internal thick cardboard structure, which was uniformly filled with a variable amount of cotton wool, plasticine, and lead shots so that it had a definite weight. Six square pieces of the corresponding surface material were glued on the faces of the cardboard structure. Because tiny slits appeared between the six pieces of the surface material, we attached white scotch tape strips on the sides of the boxes to cover the slits. In the case of the solid clay boxes, the scotch tape strips were invisible because they were painted black (see Fig. 1).

Design There were two experimental conditions, which we called *expected-weight condition* and *perceived-weight condition*. The two conditions were further divided into two separate sessions lasting 30–40 min each. Overall, each participant took part in four experimental sessions, which took place at least two days apart. All the participants first completed the two experimental sessions relative to the expected-weight condition and then the two sessions relative to the perceived-weight condition.

In both conditions, the participants were presented with *pairs* of boxes (stimuli), and they were asked to indicate which of the two stimuli in each pair appeared to be heavier or whether they appeared to weigh the same. In the expected-

weight condition, the comparative judgment had to be based only on the visual appearance of the stimuli; that is, touching the stimuli was not permitted. This was meant to quantify the contributions of surface material and size to the *expected* weight of the stimuli (see also Buckingham & Goodale, 2013; Buckingham & MacDonald, 2016; Dijker, 2008; Ellis & Lederman, 1998; Harshfield & DeHardt, 1970; Walker et al., 2010). In the perceived-weight condition, the participants used both hands together to lift each weight in turn. We could thus quantify the contributions of surface material and size to the *perceived* weight of the stimuli.

In both conditions, the participants were repeatedly presented with 30 pairs of stimuli. These pairs, which are represented by the black cells in Fig. 2, were conveniently selected among the set of all the possible 120 nonordered pairs of the 15 experimental stimuli ($15 \times 14/2 + 15$). Pilot trials showed that, when two boxes with the same surface material but a different size were compared in the expected- or the perceived-weight judgment condition, the larger box was always expected to be heavier, and it was always perceived to be lighter than the smaller box. Similarly, for two boxes with the same size but a different surface material, the box with the denser surface material was always expected to be heavier, and it was always perceived to be lighter than the box with the less dense surface material. These preliminary results were fully consistent with the basic features of the SWI and the MWI, and they indicated that the comparisons between two boxes differing in one single variable (i.e., surface material or size) would produce obvious outcomes. We thus excluded these kinds of comparisons, corresponding to the dark gray cells in Fig. 2, from the final experimental design. For similar reasons, we also excluded the comparisons in which one stimulus had both a larger size and a denser surface material than the other stimulus (the light gray cells in Fig. 2). Finally, we excluded self-comparisons (the striped cells in Fig. 2). In sum, we included in the final experimental design only the 30 pairs of stimuli characterized by a “cue conflict,” meaning that the stimulus in the pair that had the larger size had the less dense surface material (black cells in Fig. 2).

In each of the two conditions, the participants were randomly presented with four repetitions of the 30 nonordered pairs of stimuli thus selected. For each of the two conditions,

Fig. 1 **a** A picture of the 15 experimental stimuli. From front to back, the surface materials are polystyrene, wood, and clay. **b** A detailed view of the three 20-cm stimuli. From left to right, the surface materials are polystyrene, wood, and clay

		POLYSTYRENE					WOOD					CLAY				
		14	17	20	23	26	14	17	20	23	26	14	17	20	23	26
POLYSTYRENE	14															
	17															
	20															
	23															
	26															
WOOD	14															
	17															
	20															
	23															
	26															
CLAY	14															
	17															
	20															
	23															
	26															

Fig. 2 A representation of all of the 120 possible nonordered pairs of the 15 experimental stimuli. The black cells correspond to the 30 pairs of stimuli that were included in the experimental design. The striped, light

gray, and dark gray cells correspond to the pairs of stimuli that were excluded from the experimental design (see the text for explanations)

the resulting 120 trials were split into two experimental sessions of 60 trials each, which lasted 30–40 min and took place at least two days apart. The left–right order of the stimuli in each pair was counterbalanced across repetitions. All of the participants first completed the two sessions relative to the expected-weight condition and then the two sessions relative to the perceived-weight condition. We note that our methods differ from the original version of the RCM paradigm (Ho et al., 2008) in two respects. First, we presented the participants with a subset of all the possible pairs of stimuli, whereas in the original version of the paradigm, the participants were presented with all possible nonordered pairs of the (visual) stimuli. Second, rather than using a classical two-alternative forced choice task, we used a “two-alternative nonforced choice” task, which means that the indifference response was always allowed. Factors that make the latter kind of task preferable to the former are indicated by, for example, García-Pérez and Alcalá-Quintana (2011, p. 2349).

Procedure In each experimental session, the participants were seated at a square table (height 85 cm) with their elbows leaning on it. Before starting the first experimental session, they read and signed an informed consent form approved by the local ethics committee (Department of General Psychology, University of Padua). The experimenter sat in front of the participant. A screen was located approximately in the middle

of the table, so that the participants could not see the positioning and removal of the stimuli, which were manually performed by the experimenter. The stimuli were placed on a thin layer of soft rubber in order to muffle the noise of the contact of the boxes with the hard surface of the table. Once the stimuli were set into place by the experimenter, the occluding screen was removed and the participants were allowed to evaluate the weights of the stimuli either visually or both visually and haptically, as we discuss below. After the participants had responded, the experimenter reset the occluding screen in its position, removed the previous pair of stimuli from the table, and placed a new pair of stimuli on the table.

At the beginning of the two sessions of the expected weight condition, written instructions informed the participants that they would be presented with pairs of cubic boxes differing in size and material, and that their task was to indicate which of the two boxes could be expected to be heavier or if they presumably had the same weight. The instructions further specified that the boxes could not be touched in any moment during the experiment, and that the participants had to base their judgment only on the visual appearance of the boxes. Finally, the instructions specified that the material of the stimuli could be polystyrene, wood, or black-painted solid clay.

At the beginning of the two sessions of the perceived-weight condition, written instructions informed the participants that they would be presented with pairs of cubic boxes

differing in size and material and that their task was to lift the two boxes by using both hands together to lift each weight in turn, and to indicate which of the two boxes appeared to be heavier, or if they appeared to have the same weight. The instructions specified that the lifting action had to be performed by grasping each box on opposite sides and that the box on the left always had to be lifted first. The participants were further instructed to keep their elbows on the table during the lifting action, not to shake the boxes, and to respond only after the second lifting action had been completed.

Additive conjoint measurement model

We obtained estimates of the contributions of surface material and size to the expected and the perceived weights of the stimuli by fitting an *additive* conjoint measurement model to the individual data (see Ho et al., 2008). To define the model, let us denote by m_1, m_2, m_3 the three levels of the surface material variable (polystyrene, wood, and clay), and by $s_1, \dots, s_5$ the five levels of the size variable (14, 17, 20, 23, and 26 cm in side of the cubic boxes). Then, each stimulus object may be described as a pair (m_i, s_j) of a level m_i in surface material and a level s_j in size, and simply denoted by t_{ij} . For example, $t_{1,4}$ is the symbol of the 23-cm polystyrene box.

The relevant psychological property was *expected weight* in the expected-weight condition (only visual information available), and *perceived weight* in the perceived-weight condition (visual and haptic information available). In both cases, consistently with the basics of the additive conjoint measurement paradigm, we presumed that the surface material and size of a stimulus would contribute to the psychological property in the *additive* form—that is, for any stimulus t_{ij} , the combined contribution would be the sum $\mu(m_i) + \sigma(s_j)$, where $\mu(m_i)$ and $\sigma(s_j)$ are (as yet unknown) scale values associated with level m_i of the surface material and level s_j of size. Our choice of the additive response model was somewhat arbitrary, because we could equivalently use, for instance, a multiplicative response model. However, this arbitrary choice does not have any substantial implications for the outcomes of the study.

In each trial of the experiment, two stimuli, t_{ij} and t_{kl} , were presented in a two-alternative nonforced choice task. Our model presumes that the response in the trial was guided by the following rule:

t_{ij} is judged lighter, heavier, or indifferent in weight from t_{kl} ,
depending on whether
 $\mu(m_i) + \sigma(s_j) - \mu(m_k) - \sigma(s_l) + Z$ is smaller than $-\theta$, larger than θ ,
or lies between $-\theta$ and θ ,

where Z is a noise random variable with standard normal distribution, and $\theta > 0$ is a (as yet unknown) decision criterion.

In other words, we presume that in comparing stimulus t_{ij} with stimulus t_{kl} , the corresponding additive measures $\mu(m_i) + \sigma(s_j)$ and $\mu(m_k) + \sigma(s_l)$ are taken into account, and t_{ij} is judged to be lighter than t_{kl} if the difference $[\mu(m_i) + \sigma(s_j)] - [\mu(m_k) + \sigma(s_l)]$, perturbed by noise variable Z , falls below a negative cut-point $-\theta$, heavier if it falls above the symmetric positive cut-point θ , and indifferent if the difference lies between the two cut-points. This response rule is itself consistent with the random version of additive conjoint measurement (Falmagne, 1976; Ho et al., 2008). Note that in the rule, no mention is made of stimulus property w —that is, the physical weight of the boxes. This is because, in the expected-weight condition, this property was not made available to the participants, and it was constant over the whole set of stimuli in the perceived-weight condition.

The additive model thus implies a set $[\mu(m_1), \mu(m_2), \mu(m_3), \sigma(s_1), \dots, \sigma(s_5), \theta]$ of $3 + 5 + 1 = 9$ unknown parameters, which may take different values in different experimental conditions (expected vs. perceived weight) and for different participants, and which we estimated by the method of maximum likelihood. Besides the estimates, the method also renders the statistic $-\log(\text{max likelihood})$, which is a measure of the lack of fit: the greater the statistic, the worse the fit of the model to the data. This statistic is useful for judging the empirical suitability of the model by comparing it with analogous statistics implied for the same data by alternative models.

Saturated and single-factor models

For the sake of comparison, besides the additive model now described, we considered three other models to be tested on the same experimental data. One is the *saturated model*—that is, a model having as many parameters as the combinations of levels in our experimental design (plus the decision criterion parameter). When referred to the comparison between a stimulus t_{ij} and a stimulus t_{kl} , the model presumes the following response rule:

t_{ij} is judged lighter, heavier, or indifferent in weight from t_{kl} ,
depending on whether
 $\tau_{ij} - \tau_{kl} + Z$ is smaller than $-\theta$, larger than θ , or lies between $-\theta$ and θ ,

where the parameter τ_{ij} (for $i = 1, 2, 3$ and $j = 1, \dots, 5$) represents the joint contributions of surface material m_i and physical size s_j to the relevant psychological property of the stimulus t_{ij} , and Z and θ have the meanings described above. Because of its large number of parameters ($3 \times 5 + 1 = 16$), this model was expected to attain optimum fitting to the data—that is, a minimum $-\log(\text{max likelihood})$.

The other two models separately express the hypotheses that the relevant psychological property (i.e., expected or perceived weight) only depends on one of the stimulus factors in the experiment. Specifically, the *material-only*

model involves $3 + 1 = 4$ free parameters and presumes the following response rule:

t_{ij} is judged lighter, heavier, or indifferent in weight from t_{kl} ,
depending on whether
 $\mu(m_i) - \mu(m_k) + Z$ is smaller than $-\theta$, larger than θ , or lies between $-\theta$ and θ ,

where $\mu(m_i)$ and $\mu(m_k)$ are the contributions of surface materials m_i and m_k to the relevant psychological property of stimuli t_{ij} and t_{kl} . Similarly, the *size-only model* involves $5 + 1 = 6$ free parameters and presumes the following response rule:

t_{ij} is judged lighter, heavier, or indifferent in weight from t_{kl} ,
depending on whether
 $\sigma(s_j) - \sigma(s_l) + Z$ is smaller than $-\theta$, larger than θ , or lies between $-\theta$ and θ ,

where $\sigma(s_j)$ and $\sigma(s_l)$ are the contributions of sizes s_j and s_l to the relevant psychological property of stimuli t_{ij} and t_{kl} . These two models are quite simple (small number of parameters), thus they were expected to attain a worse fit to the data than would the additive model.

The four models form a three-level hierarchy, since the material-only and size-only models are nested within the additive model, which in turn is nested within the saturated model. The results we obtained in comparing these models will be commented on at the end of the next section.

Results and discussion

Table 1 shows the individual parameter estimates for the additive model in the expected-weight condition (part A) and in the perceived-weight condition (part B). Actually, the values in the tables are the differences $\mu(m_i) - \mu(m_1)$ for $i = 1, 2, 3$ and the differences $\sigma(s_j) - \sigma(s_1)$ for $j = 1, \dots, 5$, because such transformations allowed us to simplify the comparison between the scale values. These transformations are permitted because, within the frame of the RCM paradigm, the parameter estimates are determined up to linear transformations. The solid lines in Fig. 3 show the parameter estimates reported in Table 1, as they were averaged across the ten participants. Specifically, the upper panels show the averaged contributions of the three levels of variable surface material (upper left panel), and of the five levels of variable size (upper right panel), to the *expected* weights of the stimuli. The lower panels show the analogous averaged contributions to the *perceived* weights of the stimuli. Note that in Fig. 3 and hereafter in the text, subscript ε or π refers to the parameter estimates in the expected- (or perceived-) weight condition. The comparison between the upper and lower panels in Fig. 3 shows that, consistent with Prediction 1 from the expectation model, the expected weight of the stimuli increased with the density of their surface material and with their size, whereas their perceived weight decreased with both factors.

Figure 3 also shows that, inconsistently with Prediction 2, the relative contributions of surface material and size to the expected

weight of the stimuli were very different from their relative contributions to the perceived weight of the stimuli. In support of this observation, we note that the difference between the contributions of the densest and the least dense surface material to the expected weight of the stimuli [$\mu_\varepsilon(m_3) - \mu_\varepsilon(m_1)$] was about twice as big as the difference between the contributions of the largest and the smallest sizes to the expected weight [$\sigma_\varepsilon(s_5) - \sigma_\varepsilon(s_1)$]. A paired-sample t test showed that the former difference was significantly larger than the latter, as $t(9) = 7.157$, $p < .001$, Cohen's $d_z = 2.26$. In contrast, the difference between the contributions of the densest and the least dense surface material to the perceived weight of the stimuli [$\mu_\pi(m_3) - \mu_\pi(m_1)$] was about 2.5 times as small in absolute value as the difference between the contributions of the largest and the smallest size to the perceived weight [$\sigma_\pi(s_5) - \sigma_\pi(s_1)$]. A paired-sample t test showed that the latter absolute difference was significantly larger than the former one, $t(9) = 5.821$, $p < .001$, Cohen's $d_z = 1.84$.

A possible interpretation of these results, which are clearly at odds with Prediction 2, is that in the expected-weight condition the participants overestimated the contribution of surface material relative to that of size. In order to test this hypothesis, we computed the contributions of the two experimental factors to the *natural* weight of the stimuli, which we determined by emptying the stimulus boxes and weighing their shells. In other words, the natural weight is the weight that a box would have had if it were not filled so as to have the same weight as the other boxes. Let n_{ij} be the natural weight of the stimulus, characterized by the i th level of factor Surface Material and the j th level of factor Size. Also let $\mu_n(m_i)$ and $\sigma_n(s_j)$ be the contributions of the i th level of factor Surface Material, and of the j th level of factor Size, to the natural weight. For each $i = 1, 2, 3$, $\mu_n(m_i)$ was obtained by averaging the natural weights of the five stimuli with the i th level of factor Surface Material. For each $j = 1, \dots, 5$, $\sigma_n(s_j)$ was obtained by averaging the natural weights of the three stimuli with the j th level of factor Size [i.e., for each $i = 1, 2, 3$ and $j = 1, \dots, 5$, $\mu_n(m_i) = (\sum_j n_{ij})/5$ and $\sigma_n(s_j) = (\sum_i n_{ij})/3$]. For the sake of comparison with Fig. 3, Fig. 4 represents the differences $\mu_n(m_i) - \mu_n(m_1)$ (left panel) and the differences $\sigma_n(s_j) - \sigma_n(s_1)$ (right panel). The results clearly show that surface material contributed more than size to the natural weights of the stimuli. The difference between the contributions of the densest and the least dense surface materials to the natural weight [$\mu_n(m_3) - \mu_n(m_1)$] was about twice as big as the difference between the contributions of the largest and the smallest sizes to the natural weight [$\sigma_n(s_5) - \sigma_n(s_1)$]. These results show that the relative contributions of surface material and size to the *expected* weight were remarkably similar to the contributions of the two factors to the *natural* weight (see the upper panels in Fig. 3). Therefore, the violation of Prediction 2 cannot be due to unrealistic weight expectations. We will discuss possible alternative interpretations of the violation of this prediction below.

For each pair of stimuli that we presented to the participants, Fig. 5a shows the probabilities that the stimulus represented in

Table 1 Individual parameter estimates for the additive model in the expected-weight condition of Experiment 1 (A), the perceived-weight condition of Experiment 1 (B), and the perceived-weight condition of Experiment 2 (C)

(A) Experiment 1: Expected weight judgment condition									
Participant	$\mu(m_1)$	$\mu(m_2)$	$\mu(m_3)$	$\sigma(s_1)$	$\sigma(s_2)$	$\sigma(s_3)$	$\sigma(s_4)$	$\sigma(s_5)$	θ
1	0.00	0.54	2.00	0.00	0.64	1.14	1.40	2.22	0.28
2	0.00	4.51	5.11	0.00	0.53	1.68	2.48	2.75	0.97
3	0.00	4.98	7.75	0.00	0.54	1.25	1.83	2.38	1.08
4	0.00	3.30	7.75	0.00	0.51	1.23	2.40	3.09	1.04
5	0.00	4.27	6.59	0.00	1.40	2.01	3.52	2.98	0.00
6	0.00	4.70	8.22	0.00	0.94	2.65	3.49	4.98	0.89
7	0.00	4.13	8.84	0.00	1.72	2.60	4.07	5.11	0.32
8	0.00	4.31	8.38	0.00	1.46	2.79	3.37	4.51	0.48
9	0.00	3.79	7.77	0.00	1.38	2.32	3.20	4.08	1.38
10	0.00	4.41	7.76	0.00	1.24	1.71	2.99	3.25	0.78
Average	0.00	3.89	7.02	0.00	1.04	1.94	2.88	3.54	0.72
St. Error	0.00	0.40	0.65	0.00	0.14	0.20	0.26	0.33	0.14
(B) Experiment 1: Perceived-weight judgment condition									
Participant	$\mu(m_1)$	$\mu(m_2)$	$\mu(m_3)$	$\sigma(s_1)$	$\sigma(s_2)$	$\sigma(s_3)$	$\sigma(s_4)$	$\sigma(s_5)$	θ
1	0.00	-0.52	-1.89	0.00	-0.91	-1.41	-2.35	-2.86	0.29
2	0.00	-0.48	-1.00	0.00	-0.28	-0.84	-1.39	-1.45	0.47
3	0.00	-0.45	-1.40	0.00	-0.76	-1.81	-2.38	-2.85	0.72
4	0.00	-0.17	-1.25	0.00	-0.88	-1.62	-2.35	-2.40	0.72
5	0.00	-0.60	-0.92	0.00	-1.12	-2.31	-3.47	-4.25	0.82
6	0.00	-0.33	-1.11	0.00	-0.80	-1.62	-2.07	-2.81	0.44
7	0.00	-1.08	-0.75	0.00	-1.30	-1.89	-2.60	-2.79	0.00
8	0.00	-1.09	-2.05	0.00	-1.46	-2.74	-3.91	-5.21	0.04
9	0.00	-0.67	-1.16	0.00	-1.36	-2.58	-3.33	-4.10	0.95
10	0.00	-0.23	-1.15	0.00	-0.74	-1.43	-1.97	-2.34	0.61
Average	0.00	-0.56	-1.27	0.00	-0.96	-1.83	-2.58	-3.11	0.51
St. Error	0.00	0.10	0.13	0.00	0.11	0.18	0.24	0.35	0.10
(C) Experiment 2: Perceived-weight judgment condition									
Participant	$\mu(m_1)$	$\mu(m_2)$	$\mu(m_3)$	$\sigma(s_1)$	$\sigma(s_2)$	$\sigma(s_3)$	$\sigma(s_4)$	$\sigma(s_5)$	θ
1	0.00	-1.00	-1.20	0.00	-1.53	-2.79	-2.81	-4.15	0.36
2	0.00	-0.67	-0.91	0.00	-1.03	-2.00	-2.57	-3.53	0.13
3	0.00	-0.34	-0.88	0.00	-0.65	-1.31	-1.74	-2.60	0.49
4	0.00	-0.93	-0.99	0.00	-0.91	-2.03	-2.75	-3.06	0.30
5	0.00	0.15	0.04	0.00	-0.65	-1.54	-1.74	-2.41	0.45
6	0.00	-0.29	-0.30	0.00	-1.02	-1.77	-2.08	-2.82	0.27
7	0.00	-0.80	-1.15	0.00	-0.88	-2.38	-3.22	-3.53	0.16
8	0.00	0.37	0.41	0.00	-0.50	-1.03	-1.68	-1.91	0.00
9	0.00	-0.35	-0.75	0.00	-0.46	-1.22	-1.69	-2.37	0.48
10	0.00	-0.08	0.13	0.00	-0.88	-1.60	-2.36	-2.58	1.13
Average	0.00	-0.39	-0.56	0.00	-0.85	-1.77	-2.26	-2.89	0.37
St. Error	0.00	0.15	0.18	0.00	0.10	0.17	0.18	0.21	0.10

the row was *expected* to be heavier than the stimulus represented in the column. Symmetrically, Fig. 5b shows the probabilities that the stimulus represented in the row was *perceived* to be heavier than the stimulus represented in the column. On each trial, the participants' responses were assigned the values 1, 0, and .5, depending on whether the stimulus in the row was

judged to be heavier, lighter, or as heavy as the stimulus in the column. The probability values reported in Fig. 5a and b were obtained by averaging the participants' responses across the four repetitions of each pair of stimuli. Therefore, in Fig. 5a (or Fig. 5b), a probability larger than .5 indicates that the stimulus in the row tended to be expected (or perceived) to be

Fig. 3 Solid lines indicate the mean estimates of the material parameters (left panels) and of the size parameters (right panels) in the expected-weight condition (top panels) and the perceived-weight condition (bottom panels) of Experiment 1. Dashed lines (bottom panels) indicate the mean estimates of the material parameters (left panels) and of the size parameters (right panels) in the perceived-weight condition of Experiment 2. The material parameters and size parameters are measures of the contributions

of the two stimulus factors (surface material and size) to expected weight and perceived weight (see the text). The symbols m_1, m_2 , and m_3 represent the three levels of the surface material variable (polystyrene, wood, or clay), whereas the symbols $s_1, \dots, s_5$ represent the five levels of the size variable (14, 17, 20, 23, or 26 cm to a side of the cubic boxes). The vertical bars represent the standard errors of the means

heavier than the stimulus in the column. The comparison between Fig. 5a and b reveals a surprising pattern of results, which is sharply at odds with Prediction 3 from the expectation model. With a few exceptions, the stimulus in the pair that tended to be expected to be heavier also tended to be perceived to be heavier. For instance, the 14-cm wooden box was both expected and perceived to be heavier than the 23-cm polystyrene box, with probabilities .75 and .94, respectively. We marked in boldface type the probability values corresponding to the few pairs of stimuli (eight) in which the stimulus that was expected to be heavier was perceived to be lighter.

A meaningful relationship does appear between the results in Fig. 3 (estimated scale values) and those in Fig. 5a and b

(estimated probabilities). The upper panels in Fig. 3 show that the expected weights of the stimuli were mainly determined by the densities of their surface materials in a positive way. Indeed, as is shown in Fig. 5a, the stimulus in the pair with the denser surface material tended to be expected to be heavier than the stimulus with the less dense surface material, even if the former stimulus was smaller than the latter (e.g., the 14-cm clay box was expected to be heavier than the 26-cm polystyrene box). Instead, the lower panels in Fig. 3 show that the perceived weights of the stimuli were determined mainly by their sizes in a negative way. Indeed, as is shown in Fig. 5b, the stimulus in the pair with the smaller size tended to be perceived to be heavier than the stimulus with the larger size, even if the surface material of the former

Fig. 4 Estimates of the contributions of the three levels of the Surface Material factor (left panel) and of the five levels of the Size factor (right panel) to the natural weights of the stimuli. The symbols m_1, m_2 , and m_3

represent polystyrene, wood, and clay, respectively, whereas the symbols $s_1, \dots, s_5$ represent 14, 17, 20, 23, and 26 cm to a side of the cubic boxes

Fig. 5 **a** In each cell, the probability (estimated from the data of Exp. 1) shows the likelihood that the stimulus in the row was *expected* to be heavier than the stimulus in the column. **b** Probabilities (estimated from the data of Exp. 1) that the stimulus in the row was *perceived* to be heavier than the stimulus in the column. **c** Probabilities (estimated from the data of Exp. 2) that the stimulus in the row was *perceived* to be heavier than the

stimulus in the column. For panels **a** and **b**, the boldface numbers are those for the pairs of stimuli in Experiment 1 for which the stimulus expected to be heavier was also that perceived to be lighter. For panels **a** and **c**, the underlined numbers are those for the pairs of stimuli for which the stimulus *expected* to be heavier in Experiment 1 was *perceived* to be lighter in Experiment 2

was denser than the surface material of the latter (e.g., the 14-cm clay box was perceived to be heavier than the 26-cm polystyrene box). We may say that, for the set of stimuli that we used, the SWI tended to prevail over the MWI, because with few exceptions the participants' responses in the perceived-weight condition proved to depend on differences between the sizes of the stimuli, rather than on differences between the densities of their surface materials.

We conclude this section with some comments on the results we obtained in comparing the additive model with the saturated and single-factor models. The hypothesis that a model with fewer parameters performs worse than a model with more parameters can be tested by the likelihood-ratio criterion (Dobson & Barnett, 2008, p. 80), which is based on the computation of the log-likelihood-ratio statistic (LLR). Because we were interested in a general comparison between the four models (i.e., saturated, additive, size-only, and material-only), we fitted them to two pooled datasets, which are the set of data from the ten participants in the expected-weight judgment condition and the analogous set in the perceived-weight judgment condition.³ The essential statistics that we obtained are shown in Table 2, in which, besides the numbers of free parameters in the models, we give the $-\log(\text{max likelihood})$ statistic resulting in each fit of the model to the data. The LLR statistics for the comparison between the additive model (i.e., the nested model) and the saturated

model (the reference model) in the expected-weight and perceived-weight conditions were 1.68 and 9.7, respectively, both below the critical point $\chi^2(.05, 16^{-9}) = 14.067$. This signifies that the additive model should be preferred to the saturated model because the former does not perform significantly worse than the latter in terms of goodness of fit, and it is simpler in terms of the number of parameters. The LLR statistics for comparing the size-only versus the additive model were 686.76 and 134.26, which fall well above $\chi^2(.05, 9^{-6}) = 7.814$; the statistics for comparing the material-only versus the additive model were 149.82 and 412.8, which are definitely larger than $\chi^2(.05, 9^{-4}) = 11.070$. We may then conclude that neither the size-only nor the material-only model is suitable to replace the additive model, because of their much worse fits to the data.

Experiment 2

So far, we have focused only on the contrast between actual and expected weight as a possible cause of weight illusions. However, Amazeen and Turvey (1996) showed that the negative relationship between size and perceived weight (i.e., the

Table 2 For each model that we fitted to the pooled data of Experiment 1, the table shows the number of parameters (N Par.) and the values of the $-\log(\text{max likelihood})$ statistic for the two experimental conditions

Model	N Par.	$-\log(\text{max likelihood})$ Expected Weight	$-\log(\text{max likelihood})$ Perceived Weight
Saturated	16	599.41	1,031.81
Additive	9	600.25	1,036.67
Size-only	6	943.62	1,103.80
Material-only	4	675.15	1,243.07

³ When dealing with the individual data, we found that for some participants the additive model fitted slightly better than the saturated model, which indicates that for the saturated model the searching procedure failed to converge to the optimal solution (maximum likelihood). This is probably due to the fact that the number of parameters of the saturated model was too high, as compared with the number of trials for each participant. We decided to bypass this difficulty by fitting the four models to the pooled data, which provide an overall picture of the performance of the models.

SWI) may also depend on people's sensitivity to the rotational inertia of objects. Rotational inertia is a higher-order physical variable that refers to the resistance of objects to rotation in space, and it depends on a complex interaction between the masses, the sizes, and the shapes of objects (i.e., it depends on the mass distribution of the objects). The effects of rotational inertia on perceived weight appear to depend both on the mode of lifting and on the physical properties of the lifted objects. For instance, the effects of rotational inertia were found to be strong in the case of elongated objects that were rotated in space (Amazeen, 1997, Exp. 1; Amazeen & Turvey, 1996), but to be negligible in the case of spherical objects that were grasped and held in the hand (Zhu, Shockley, Riley, Tolston, & Bingham, 2013).

A tentative interpretation of the results of Experiment 1 is that the perceived weights of the stimuli depended not only on the contrast between the actual and expected weights (i.e., a top-down perceptual process), but also on participants' sensitivities to rotational inertia (i.e., a bottom-up perceptual process). When an object is grasped and lifted, as in Experiment 1, weight expectations and rotational inertia are both available to the participants, because the former are prompted by the visual appearance of the stimuli (Buckingham & Goodale, 2010), whereas the latter can be "picked up" by the participants through the lifting action (Amazeen & Turvey, 1996). If weight expectations and rotational inertia jointly contributed to the perceived weights of the stimuli, then the negative relationship between the *size* of the stimuli and their perceived weight would depend upon two distinct and independent perceptual processes: first, the contrast between actual and expected weights, and second, the participants' sensitivities to rotational inertia. Instead, because rotational inertia is independent of surface material, the negative relationship between the *density* of the surface material of the stimuli and their perceived weight would only depend upon the contrast between the actual and expected weights. According to this hypothesis, the negative relationship between size and perceived weight should be stronger than that between the density of the surface material and perceived weight, because the former relationship would subtend two distinct perceptual processes rather than one. The results of Experiment 1 are consistent with this hypothesis, which might provide a suitable explanation for the violations of Predictions 2 and 3 from the expectation model. We designed Experiment 2 to provide an empirical test of this hypothesis.

In a different manner from Experiment 1, the participants in Experiment 2 were asked to lift the stimuli by using a string that was attached to the top surface, because this procedure should eliminate or greatly reduce the perceptual information available regarding rotational inertia (Amazeen & Turvey, 1996). If in Experiment 1 the violation of the predictions from the expectation model was due to the bottom-up influence of rotation inertia on perceived weight, then those predictions

should be met in Experiment 2, in which rotational inertia could not influence the perceived weights of the stimuli. Moreover, if rotational inertia had affected the perceived weights of the stimuli in Experiment 1, then the relative contribution of size to perceived weight should be smaller in Experiment 2 than in Experiment 1.

Method

Participants Ten participants (five females, five males) took part in the experiment. They were undergraduate or graduate students at the Faculty of Psychology, University of Padua, and ranged in age from 20 to 26 years ($M = 22.4$, $SD = 1.96$). All had normal or corrected-to-normal vision and normal function of their upper limbs. They were naive as to the purposes of the experiment and received €10 for their participation. None of them had participated in Experiment 1.

Stimuli The stimuli were the same as in Experiment 1, except that a strong 20-cm string was attached to the top surface of each box.

Design The design of Experiment 2 was the same as that of Experiment 1, except that in the former we tested only the perceived-weight condition. Because the mode of lifting should not affect the expected weight of the stimuli, we compared the perceived weights of the stimuli in Experiment 2 with their expected weights in Experiment 1.

Procedure The procedure was the same as in the perceived-weight condition of Experiment 1, except for the following details. The participants stood in front of the table and were asked to use their favored hand. They were instructed to lift the stimuli by grasping the string at the top end, and to move their forearm so that the stimuli were raised and lowered vertically, without any lateral motion or rotation.

Results and discussion

The dashed lines in the lower panels of Fig. 3 show the parameter estimates for the additive model that are reported in Table 1 (part C), as they were averaged across the ten participants. The results show that the relative contribution of size to the perceived weight of the stimuli was decidedly larger in absolute value than that of surface material, which means that Prediction 2 was also violated in Experiment 2. Moreover, Fig. 5c shows the probabilities that the stimulus in the row was perceived to be heavier than the stimulus in the column. The comparison between Fig. 5a and c reveals several violations of Prediction 3, in that, with only two exceptions, the stimulus in the pair that tended to be expected to be heavier (in Exp. 1) also tended to be perceived to be heavier (in Exp. 2). We underlined the probability values corresponding to the two

pairs of stimuli in which, consistent with Prediction 3, the stimulus that was expected to be heavier was perceived to be lighter. In sum, clear violations of Predictions 2 and 3 also appeared in Experiment 2, which disconfirms the hypothesis that the violation of such predictions could be imputed to the bottom-up effects of rotational inertia on perceived weight.

Though the bottom-up influence of rotational inertia on perceived weight cannot explain the violation of the predictions from the expectation model, it still might have exerted some influence on the perceived weight of the stimuli in Experiment 1. In order to test this hypothesis, we compared the relative contributions of surface material and size in the two experiments by means of the measure $[\sigma_{\pi}(s_5) - \sigma_{\pi}(s_1)] / [\mu_{\pi}(m_3) - \mu_{\pi}(m_1)]$. We computed this measure on the individual data of both experiments: The larger the measure, the larger the contribution of size to the perceived weight, as compared with the contribution of surface material. An independent-sample t test showed that the measure in Experiment 2 was not significantly different from the measure in Experiment 1, since $t(9.04) = 0.6$, $p = .56$, Cohen's $d_s = 0.27$. A Bayesian independent-sample t test provided support for the null hypothesis of equivalence between the two experiments in that measure, since the JZS $BF_{01} = 2.21$. We thus conclude that rotational inertia exerted a null or negligible influence on the perceived weights of the stimuli in our study. Zhu et al. (2013) obtained a similar result for a set of spherical objects differing in size, mass, and mass distribution. As regards the statistical comparison of the fits of the models to the data, the results were in line with those of Experiment 1: The additive model performed significantly better than the single-factor models, and it did not perform significantly worse than the saturated model.

General discussion

The results of the present study showed some systematic violations of the predictions from the expectation model. These violations proved to be consistent across different modes of lifting (i.e., grasping the stimuli on opposite sides, or lifting them by means of a string attached to their top surface). Inconsistent with Prediction 2 from the model, the *expected* weight depended more on the surface materials than on the sizes of the stimuli, whereas the *perceived* weight depended more on their sizes than on their surface materials. The inconsistency in the relative contributions of the two factors between the two experimental conditions was also revealed by the fact that, for 22 out of the 30 pairs of stimuli in Experiment 1, and for 28 out of the 30 pairs of stimuli in Experiment 2, the stimulus *expected* to be heavier in a pair was also *perceived* to be heavier. This peculiar result is inconsistent with Prediction 3, and is especially at odds with the idea that the perceived weights of objects can simply be explained in terms

of the contrast between their physical and expected weights. Buckingham and MacDonald (2016) recently reported a similar finding, and the results of our study extend and generalize this finding to a large set of objects that jointly differed in two properties—that is, surface material and size.

For technical reasons, the stimuli in our experiments were quite heavy (i.e., 1.2 kg), as compared with those typically employed in most weight illusion studies. Could the outcomes of our study have been somehow dependent on this feature of the stimuli? Weight illusions usually depend on small differences in perceived weight; thus, their magnitude is affected by weight discrimination (i.e., people's sensitivity to weight differences). It has been shown (by, e.g., Jones & Burgess, 1998; Ross & Gregory, 1970; see Nicolas et al., 2012, p. 124, for a discussion) that weight discrimination is lower when the physical weights of the stimuli exceed the range of weights within which the perceptual system is optimally adapted. In the light of this finding, we cannot exclude that the absolute magnitudes of the SWI and the MWI in our study could have been lower than has been shown in studies in which lighter stimuli were used. Nonetheless, our test of the expectation model was based on hypotheses concerning the *relative* magnitudes of the two illusions, rather than on hypotheses concerning their absolute magnitudes. Insofar as weight discrimination affects the magnitudes of both illusions in similar ways, as it appears reasonable to presume, the main outcomes of our study can be generalized to a wide range of physical weights.

As we argued above, the violations of Predictions 2 and 3 from the expectation model could not be explained in terms of unrealistic weight expectations, or in terms of the bottom-up influence of rotational inertia on perceived weight. Dijker (2008, 2014) suggested that multiple perceptual processes may contribute to the perceived weights of objects. In Dijker's (2008) study, the stimuli were toy dolls that varied in apparent size and apparent physical strength; similar to what we found in the present study, the author found that the relative contributions of the two experimental factors to the expected weight were clearly different from their relative contributions to perceived weight. Because the contrast between actual and expected weights, taken alone, could not explain the experimental results, the author hypothesized that the perceived weights of the toy dolls were also affected by the contrast between the force with which the dolls were lifted and their actual weights. However, the results of Buckingham and Goodale's (2013) study showed that, for objects varying in surface material and size—as was the case with the stimuli in our experiments—the lifting forces do not appear to exert a significant influence on perceived weight. In considering this evidence, we suggest that Dijker's (2008) explanation cannot be generalized to the results of our study. However, we add a note of caution, because the stimuli in Buckingham and Goodale's (2013) study

were lighter than those in our study, and little is known about the relationship between lifting forces and perceived weights in the case of heavy objects.

Buckingham (2014; see also Buckingham & Goodale, 2013; Buckingham & MacDonald, 2016) recently proposed a revision of the expectation model, according to which weight perception is affected by implicit, rather than explicit, weight expectations. Explicit weight expectations can be directly measured by asking participants to predict the weight of an object on the basis of its visual appearance, as we did in the present study. Implicit weight expectations would instead be impervious to consciousness and would depend on ontogenetic and phylogenetic development. Buckingham (2014) highlighted various differences between explicit and implicit weight expectations; for the purposes of our discussion, here we focus on the hypothesis that *explicit* expectations are mostly affected by the densities of the surface materials of objects, and only to a lesser extent by their sizes, whereas *implicit* expectations would mostly be affected by the sizes of objects rather than by the densities of their surface materials. Therefore, whereas the participants may *explicitly* expect that a small box with a relatively dense surface material (e.g., 14-cm clay box) would be heavier than a large box with a low-density material (e.g., a 23-cm polystyrene box), they may *implicitly* expect that the latter would be heavier than the former. If the alleged difference between implicit and explicit expectations is correct, and if it is true that perceived weight is influenced by implicit rather than explicit expectations, then the results of our study would be consistent with Buckingham's (2014) revised version of the expectation model.

Author note We thank Alice Fusaro for help in the construction of the experimental stimuli.

References

- Amazeen, E. L. (1997). The effects of volume on perceived heaviness by dynamic touch: With and without vision. *Ecological Psychology*, 9, 245–263. doi:10.1207/s15326969eco0904_1
- Amazeen, E. L., & Turvey, M. T. (1996). Weight perception and the haptic size–weight illusion are functions of the inertia tensor. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 213–232. doi:10.1037/0096-1523.22.1.213
- 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, 1518–1531. doi:10.1152/jn.00814.2009
- Buckingham, G. (2014). Getting a grip on heaviness perception: A review of weight illusions and their probable causes. *Experimental Brain Research*, 232, 1623–1629. doi:10.1007/s00221-014-3926-9
- 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, e9709. doi:10.1371/journal.pone.0009709
- Buckingham, G., & Goodale, M. A. (2013). Size matters: A single representation underlies our perceptions of heaviness in the size–weight illusion. *PLoS ONE*, 8, e54709. doi:10.1371/journal.pone.0054709
- Buckingham, G., & MacDonald, A. (2016). The weight of expectation: Implicit, rather than explicit, prior expectations drive the size–weight illusion. *Quarterly Journal of Experimental Psychology*, 69, 1831–1841. doi:10.1080/17470218.2015.1100642
- Buckingham, G., Ranger, N. S., & Goodale, M. A. (2011). The material–weight illusion induced by expectations alone. *Attention, Perception, & Psychophysics*, 73, 36–41. doi:10.3758/s13414-010-0007-4
- Charpentier, A. (1891). Analyse expérimentale de quelques éléments de la sensation de poids. *Archives de Physiologie Normale et Pathologique*, 3, 122–135.
- Davis, C. M., & Roberts, W. (1976). Lifting movements in the size–weight illusion. *Perception & Psychophysics*, 20, 33–36. doi:10.3758/BF03198701
- Dijker, A. J. M. (2008). Why Barbie feels heavier than Ken: The influence of size-based expectancies and social cues on the illusory perception of weight. *Cognition*, 106, 1109–1125. doi:10.1016/j.cognition.2007.05.009
- Dijker, A. J. M. (2014). The role of expectancies in the size–weight illusion: A review of theoretical and empirical arguments and a new explanation. *Psychonomic Bulletin & Review*, 21, 1404–1414. doi:10.3758/s13423-014-0634-1
- Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models*. Boca Raton, FL: CRC Press.
- Ellis, R. E., & Lederman, S. J. (1998). The golf-ball illusion: Evidence from top-down processing in weight perception. *Perception*, 27, 193–201. doi:10.1068/p270193
- Falmagne, J. C. (1976). Random conjoint measurement and loudness summation. *Psychological Review*, 83, 65–79. doi:10.1037/0033-295X.83.1.65
- Flanagan, J. R., & Beltzner, M. A. (2000). Independence of perceptual and sensorimotor predictions in the size–weight illusion. *Nature Neuroscience*, 3, 737–741. doi: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, 1742–1747. doi:10.1016/j.cub.2008.09.042
- García-Pérez, M. A., & Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: Sorting out the artifacts. *Attention, Perception, & Psychophysics*, 73, 2332–2352. doi:10.3758/s13414-011-0167-x
- 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. doi:10.1007/BF00229824
- Grandy, M. S., & Westwood, D. A. (2006). Opposite perceptual and sensorimotor responses to a size–weight illusion. *Journal of Neurophysiology*, 95, 3887–3892. doi:10.1152/jn.00851.2005
- Harshfield, S. P., & DeHardt, D. C. (1970). Weight judgments as a function of apparent density of objects. *Psychonomic Science*, 20, 365–366.
- Ho, Y.-X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, 19, 196–204. doi:10.1111/j.1467-9280.2008.02067.x
- Jones, L. F., & Burgess, P. R. (1998). Neural gain changes subserving perceptual acuity. *Somatosensory and Motor Research*, 15, 190–199. doi:10.1080/08990229870754
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York, NY: Springer. doi:10.1007/978-1-4614-4475-6
- Masin, S. C., & Crestoni, L. (1988). Experimental demonstration of the sensory basis of the size–weight illusion. *Perception & Psychophysics*, 44, 309–312. doi:10.3758/BF03210411
- Nakatani, K. (1985). Application of the method of fixed set to the size–weight illusion. *Psychological Research*, 47, 223–233. doi:10.1007/BF00309450
- Nicolas, S., Ross, H. E., & Murray, D. J. (2012). Charpentier's papers of 1886 and 1891 on weight perception and the size–weight illusion. *Perceptual and Motor Skills: Perception*, 115, 120–141. doi:10.2466/24.22.27.PMS.115.4.120-141

- Plaisier, M. A., & Smeets, J. B. J. (2012). Mass is all that matters in the size–weight illusion. *PLoS ONE*, 7, e42518. doi:[10.1371/journal.pone.0042518](https://doi.org/10.1371/journal.pone.0042518)
- Ross, H. E. (1969). When is a weight not illusory? *Quarterly Journal of Experimental Psychology*, 21, 346–355. doi:[10.1080/14640746908400230](https://doi.org/10.1080/14640746908400230)
- Ross, H. E., & Gregory, R. L. (1970). Weight illusions and weight discrimination—A revised hypothesis. *Quarterly Journal of Experimental Psychology*, 22, 318–328. doi:[10.1080/00335557043000267](https://doi.org/10.1080/00335557043000267)
- Seashore, C. E. (1899). Some psychological statistics: II. The material weight illusion. *University of Iowa Studies in Psychology*, 2, 36–46.
- Walker, P., Francis, B. J., & Walker, L. (2010). The brightness–weight illusion: Darker objects look heavier but feel lighter. *Experimental Psychology*, 57, 462–469. doi:[10.1027/1618-3169/a000057](https://doi.org/10.1027/1618-3169/a000057)
- Zhu, Q., Shockley, K., Riley, M. A., Tolston, M. T., & Bingham, G. P. (2013). Felt heaviness is used to perceive the affordance for throwing but rotational inertia does not affect either. *Experimental Brain Research*, 224, 221–231. doi:[10.1007/s00221-012-3301-7](https://doi.org/10.1007/s00221-012-3301-7)