The intrinsic constraint model and Fechnerian sensory scaling

Fulvio Domini

Department of Cognitive and Linguistic Sciences,
Brown University, Providence, RI, USA

Corrado Caudek

Department of Psychology,
University of Florence, Firenze, Italy

The Intrinsic Constraint (IC) model of depth-cue integration (F. Domini, C. Caudek, & H. Tassinari, 2006) posits a strong link between perceived depth and depth discrimination, much like some Fechnerian theories of sensory scaling. K. J. MacKenzie, R. F. Murray, and L. M. Wilcox (2008) tested the IC model by examining whether two depth-matched pairs of stimuli are separated by equal numbers of Just Noticeable Differences (JNDs) in depth. They concluded that "IC is inconsistent with the psychophysics of depth perception." Here, by using a different methodological approach, we provide empirical findings that are consistent with the predictions of the IC model. We also discuss the relative merits of the IC and Modified Weak Fusion (MWF) models (M. S. Landy, L. T. Maloney, E. B. Johnston, & M. Young, 1995) of depth-cue combination.

Introduction

In a recently published article, MacKenzie, Murray, and Wilcox (2008) presented an empirical test of the Intrinsic Constraint (IC) model of depth-cue integration (Domini, Caudek, & Tassinari, 2006), maintaining that "IC has similarities to Fechnerian theories of sensory scaling, in that it predicts that perceived depth can be meaningfully measured in terms of Just Noticeable Differences (JNDs)." The rationale of the investigation of MacKenzie et al. is the following. Assume that two pairs of stimuli are matched in terms of their depth differences, with each pair being defined by a different depth cue. If Fechnerian theory holds, then we should expect that the first stimulus pair would be separated by the same number of JNDs as the second. MacKenzie et al. found that Fechner's hypothesis is not supported for JNDs computed in a depth-discrimination task for motion and stereo stimuli. They concluded that "IC is inconsistent with the psychophysics of depth perception."

In the present paper, we report a new investigation in which we construct a psychophysical scale of depth perception through the cumulation of psychometric increments. Contrary to the conclusions of MacKenzie et al. (2008), our results reveal that the relation between perceived depth magnitudes and JND sums is compatible with Fechner's theory.

The rest of this paper is organized as follows. In the Stairway to depth perception section, we review the experimental design of MacKenzie et al. (2008). In the Theoretical analysis section, we provide the theoretical arguments necessary for a proper test of the hypothesis of MacKenzie et al. (2008). Finally, in An empirical test of the IC model section, we present the empirical data of an appropriate test of the IC model.

Stairway to depth perception

Before describing the experiment of MacKenzie et al. (2008), we introduce the notation that will be used throughout this paper (see also the Table 1). The superscript "(j)" will index the *j*th amount of *perceived depth,* with *j* = 1,…. The subscript "$_v$" or "$_d$" will denote, respectively, the velocity or disparity signals provided by single-cue stimulus displays. By writing $z_v^{(1)}$ and $z_d^{(1)}$, we will indicate the *simulated* depth magnitudes of a pair of velocity-only and disparity-only stimuli, which give rise to the same amount of perceived depth, here indexed by *j* = 1.

The stimulus displays used by MacKenzie et al. (2008) consisted of computer-generated random-dot patterns simulating a half-cylinder. The three-dimensional (3D) structure of the half-cylinder was specified by either motion or stereo information. In the *depth-matching* part of the experiment, observers compared half-cylinders defined by stereo information (*test*) with three half-cylinders defined by motion information (*standard*). The *standard* half-cylinders were 12.5, 25, and 50 mm deep. In a Two-Interval Forced-Choice (2-IFC) task, observers indicated which of the two successively presented half-cylinders

Notation	Meaning
^	Denotes a perceptual estimate
Δd	Disparity difference
$\Delta \hat{z}$	Increment in perceived depth
ε_1	Disturbance on ρ distributed as $\mathcal{N}(0, 1)$
$\varepsilon_{d_i}, \varepsilon_{v_i}$	Measurement errors of the disparity and velocity signals, respectively
$\varepsilon_{\hat{z}_i}$	Disturbance of perceived depth $\hat{z}_i$
μ	Vergence angle
ρ_i	Scores of the ith surface point on the first Principal Component computed from the scaled image signals
ρ_P	Intensity of ρ at point P
σ_v, σ_d	SD of measurement noise for velocity and disparity signals, respectively
$\sigma_{\hat{z}}$	SD of the perceived depth noise
σ_D	SD of judgment noise introduced in the depth interpretation stage
ω	Angle of rotation in 3D space
P	A generic surface point
d_i, v_i	Relative disparities and velocities, respectively
d_P, v_P	Relative disparities and velocities at point P
$f_\rho(\rho)$	Positive monotonic function of ρ
$f'_\rho(\rho_P)$	First-order derivative of $f_\rho(\rho)$ at ρ_P
$z_x^{(j)}$	z amount of simulated depth corresponding to the jth amount of perceived depth; the index identifies the depth cue provided by the stimulus display
z_i	Depth map where i is the index of the feature points

Table 1. Notation table.

appeared to be deeper. By varying the simulated depth of the *test* half-cylinders according to a staircase procedure, the authors estimated the Point of Subjective Equality (PSE) for perceiving the *test* (stereo-only) and *standard* (motion-only) half-cylinders with the same depth. In this part of the experiment, MacKenzie et al. (2008) found the three stereo-depths $z_d^{(1)}$, $z_d^{(2)}$, and $z_d^{(3)}$, which perceptually matched the three motion depths $z_v^{(1)}$, $z_v^{(2)}$, and $z_v^{(3)}$ of the *standard* stimuli.

In the *depth-discrimination* part of the experiment, observers were asked to perform a depth-discrimination task at each of the three simulated depth magnitudes. This task was performed separately for the motion and stereo stimuli. Through this procedure, the authors estimated three JNDs for the stereo stimuli ($JND_d^{(1)}$, $JND_d^{(2)}$, and $JND_d^{(3)}$) and three JNDs for the motion stimuli ($JND_v^{(1)}$, $JND_v^{(2)}$, and $JND_v^{(3)}$). The authors then asked the critical question: How many JNDs separate the $z_v^{(1)}$ and $z_v^{(3)}$ motion-only stimuli, on the one hand, and the perceptually matched $z_d^{(1)}$ and $z_d^{(3)}$ stereo-only stimuli, on the other? If Fechner's theory holds, then the motion-only stimuli should be separated by the same number of JNDs as the perceptually matched stereo-only stimuli. MacKenzie et al. (2008) found that the "depth-matched pairs of stimuli were not (…) separated by equal numbers of JNDs, contradicting IC's prediction."

In order to discuss the results of MacKenzie et al., suppose that Fechner's theory holds. Suppose also that the depth value denoted by $z^{(1)}$ is separated by five JNDs from the depth value denoted by $z^{(6)}$. We can envision two scenarios. In the first scenario, the JNDs for stereo and motion stimuli are constant within the depth range examined in the experiment. In order to estimate the number of JNDs separating the smallest depth $z^{(1)}$ from the largest depth $z^{(6)}$, therefore, it is sufficient to divide the depth difference $z^{(6)} - z^{(1)}$ (i.e., the height of the scale) by the size of one JND (i.e., the height of a step); see Figure 1 (left panel). In the second scenario, the JNDs are not constant (Figure 1, right panel). To estimate the JND count, therefore, it would not be appropriate to divide the total depth difference by either the first or the last JND comprising the psychological scale.

MacKenzie et al. (2008) estimated the number of JNDs separating two disparity-defined or two motion-defined stimuli in three different ways. In their Figures 8 and 9, MacKenzie et al. (2008) report the JND count computed by dividing the simulated depth difference of each pair of stimuli by the JND from the shallower or the deeper stimulus, respectively. In a further analysis, MacKenzie et al. (2008) estimated the number of JNDs by taking into account the Weber's law. MacKenzie et al. (2008) reasoned as follows. If the JND magnitudes are related to the stimulus depth with a proportionality constant k, then the number of JNDs separating two disparity-defined or motion-defined stimuli at depths z_1 and z_2 (with $z_2 > z_1$) will be equal to $n = (1/k)(\log|z_2| - \log|z_1|)$. The JND counts estimated in this manner are reported in their Figure 10. On the basis of all these results, MacKenzie et al. (2008) concluded that depth-matched motion and disparity displays are not separated by the same number of JNDs.

The previous literature on depth-discrimination thresholds (e.g., Enright, 1991) suggest that, in order to test Fechner's hypothesis in the domain of depth perception, it is necessary to base the JND count on the scenario represented by the right panel of Figure 1. The results reported in Figures 8 and 9 of MacKenzie et al. (2008), therefore, cannot be considered as conclusive. Moreover, the assumption that depth-discrimination performance approximately follows Weber's law (Figure 10 of MacKenzie et al., 2008) is also questionable. There are many reasons to question Weber's law. Evidence in this regard comes, for example, from our previous data. The predictor ρ, which we used in our previous works, is equal to the signal-to-noise ratio in the case of single-cue displays. If depth-discrimination performance followed Weber's law, ρ would remain constant with simulated depth and our model would not fit the psychophysical data. All our previous findings concerning the IC model, however, contradict such hypothesis (Domini et al., 2006; Tassinari, Domini, & Caudek, 2008). Other evidence contrary to Weber's law comes from Farell, Li, and McKee (2004a). They found that disparity thresholds for random-dot
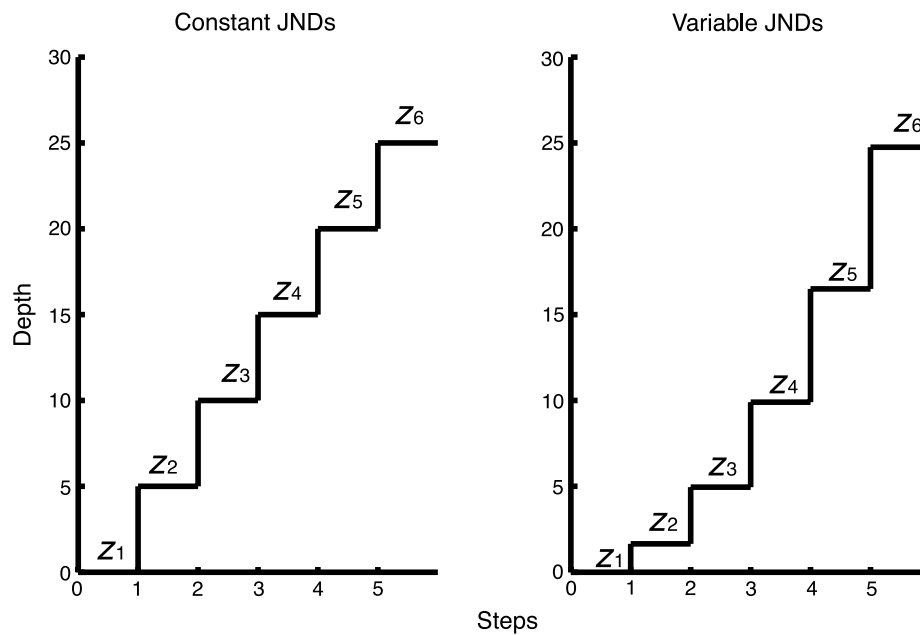
Figure 1. Representation of subjective (Fechnerian) distances. Each psychophysical scale corresponds to a cumulation of psychometric increments ("steps"). In the figure, the physical depth corresponding to a Just Noticeable Difference (JND) is shown as a function of the location on the physical continuum in which observers were asked to perform a depth-discrimination task. (Left panel) Constant JNDs. (Right panel) JNDs increasing with stimulus intensity.

stereograms increased as a function of pedestal disparity by following an exponential law with an intercept different from zero. Such result cannot be accounted for by Weber's law. On the basis of such evidence, we decided to test the hypothesis of MacKenzie et al. (2008) by using a different methodology.

Theoretical analysis

In order to relate the IC model to a JND sum, we will proceed as follows. In The intrinsic constraint model section, we will review the IC model. In the JND and measurement error section, we will discuss a critical assumption of MacKenzie et al. (2008): The assumption that the JND provides an unbiased estimate of the standard deviation of the disparity and velocity noise. We will then demonstrate that the JND is indeed a biased estimator of the disparity and velocity noise. In A proper test of the IC model section, we will show how it is possible to provide a proper test for the hypothesis of MacKenzie et al. (2008). Finally, in the IC and Fechnerian scaling section, we will discuss the relation between the IC model and Fechnerian scaling.

The intrinsic constraint model

IC is a two-stage model. The goal of the first stage is to recover a precise and accurate estimate of local affine

structure. In the Local affine structure and signal-to-noise ratio section, we will show that the precision of the estimation of local affine structure is provided by the signal-to-noise ratio. In IC: Maximum SNR estimate of local affine structure section, we will show that the weighted combination of image signals computed by the IC model has the maximum signal-to-noise ratio; this indicates that the IC model provides the best estimator of local affine structure. In the second stage, the IC model imposes a maximum-likelihood metric depth interpretation to the recovered affine structure (Tassinari et al., 2008).

Local affine structure and signal-to-noise ratio

If z_i is a depth map, with i being the index of the feature points, the 3D *affine structure* is a family of depth maps, where k can take on any value. Affine transformations preserve affine properties, such as depth-order relationships, parallelism, and so on (e.g., Koenderink & van Doorn, 1991; Todd, Oomes, Koenderink, & Kappers, 2001).

One important property of retinal projections is that binocular disparities and retinal velocities *directly specify* the local affine structure within the visual scene. For small visual angles, in fact, it can be shown that

$$d_i = \mu z_i + \varepsilon_{d_i} = \mathbb{E}(d_i) + \varepsilon_{d_i}, \tag{1}$$

$$v_i = \omega z_i + \varepsilon_{v_i} = \mathbb{E}(v_i) + \varepsilon_{v_i}, \tag{2}$$

where μ is the vergence angle and ω is the angle of rotation in 3D space; the disturbance terms ε_{d_i} and ε_{v_i}, due to measurement errors, are modeled as Gaussian noise with zero mean and standard deviations σ_{d_i} and σ_{v_i}, respectively.

In the presence of image noise, the estimates of local affine structure vary from one set of measurements to another. It is easy to recognize that the expected values of the velocity and disparity signals are proportional to the depth map: $\mathbb{E}(d_i) = \mu z_i$ and $\mathbb{E}(v_i) = \omega z_i$. In other words, velocity and disparity signals specify the local affine structure.

How can we quantify the precision with which local affine structure is estimated? Such a precision cannot be quantified in terms of the absolute value of the standard deviations of the measurement noise, σ_{d_i} and σ_{v_i}. Any linear scaling of the image signals, in fact, affects the magnitudes of σ_{d_i} and σ_{v_i} but leaves the precision of estimation unchanged. To clarify this point, consider the disparity signals as estimators of the depth-order relations. Intuitively, the precision of estimation provides a quantitative answer to the following question: "In what way do the relations among the disparity signals reflect the depth-order relations?" A mismatch between the relations among the disparity signals and the depth-order relations is indicated in Figure 2. In each of the four cases S, with $S \in \{1,2,3,4\}$, consider the depth-order relation $P_1 < P_2$, where $P_2 = P_1 + \Delta P$. The point P_2 is always in front of P_1. Because of noise, however, sometimes the disparity signal d_{P_2} is larger than d_{P_1}, but sometimes smaller. A violation of the order relations in the recovered affine structure occurs if $d_{P_1} > d_{P_2}$.

The likelihood of a depth reversal increases with (a) a decrease of the difference $\mathbb{E}(\Delta d) = \mathbb{E}(d_{P_2}) - \mathbb{E}(d_{P_1})$ and (b) an increase of the standard deviation σ_d of the disparity noise. Since we are dealing with a local analysis, without loss of generality, we assume that the standard deviations of the measurement noise are the same for both P_1 and P_2. The precision of the estimation of the local affine structure, therefore, can be quantified as

$$\text{SNR} = \frac{\mathbb{E}(\Delta d)}{\sigma_d} \qquad (3)$$

the Signal-to-Noise Ratio (SNR). The SNR can thus be used as criterion to compare the precision of the estimates derived from different depth cues.

In the previous discussion, we assumed that the noise of disparities and velocities is nearly constant across the local region of interest. Even though the demonstration is not provided here, it can be shown that the same conclusions that we have reached presently are also found in the more general case of non-constant noise.

IC: Maximum SNR estimate of local affine structure

Di Luca, Domini, and Caudek (2007) showed that the best estimator of the local affine structure is provided by a weighted combination of the image signals that maximizes the SNR of the resulting decision variable. Consider here the case of two signals, disparity and
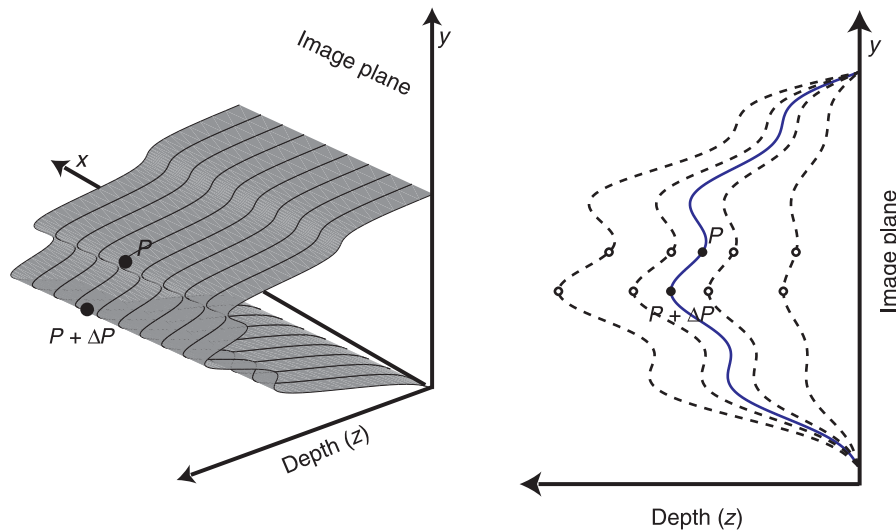

Figure 2. (Left panel) A 3D structure with a constant *y, z* profile. The *z*-axis represents the depth axis and points towards the observer. *P* and *P* + Δ*P* are two feature points belonging to the surface. (Right panel) A side view of the 3D surface (solid blue line) and four affine stretches (dashed lines) preserving affine properties. For example, all four surfaces represented by the dashed lines exhibit the same depth-order relations among the neighboring points *P* and *P* + Δ*P*. The point *P* + Δ*P* always lies in front of *P*.

velocity. As also pointed out by MacKenzie et al. (2008), the SNR is maximized by a weighted sum

$$r_i = w_{d_i} d_i + w_{v_i} v_i \qquad (4)$$

in which the weights w_{d_i} and w_{v_i} are given by

$$w_{d_i} = \frac{\mu \sigma_{d_i}^{-2}}{\sqrt{\left(\mu \sigma_{d_i}\right)^{-2} + \left(\omega \sigma_{v_i}\right)^{-2}}}, \qquad (5)$$

$$w_{v_i} = \frac{\omega \sigma_{v_i}^{-2}}{\sqrt{\left(\mu \sigma_{d_i}\right)^{-2} + \left(\omega \sigma_{v_i}\right)^{-2}}}. \qquad (6)$$

In general, the weights must be $w_d \propto \frac{E(d_i)}{\sigma_d^2}$ and $w_v \propto \frac{E(v_i)}{\sigma_v^2}$. With the particular choice of the weights in the previous equation, the output noise will have unit variance. The weighted sum r_i will thus be equal to $r_i = \rho_i + \varepsilon_r$ with $\varepsilon_r \sim \mathcal{N}(0, 1)$ and $E(r_i) = \rho_i$.

Domini et al. (2006) showed that the weights described by Equations 5 and 6 can be estimated by a Principal Component Analysis (PCA) carried out on the disparity (d_i) and velocity (v_i) signals scaled by the standard deviation of their measurement noise. The scores (r_i) on the first principal component correspond to the optimal combination of Equation 4, with the weights indicated by Equations 5 and 6. This combination is optimal in the sense that it maximizes the SNR of the combined estimate (see also Tassinari et al., 2008).

JND and measurement error

In their test of the IC model, MacKenzie et al. (2008) assumed that the JND provides an unbiased estimate of the standard deviation of the measurement error. This assumption, critical for the test of the IC model, requires a specification of the different sources of stochastic noise, which affect the task under examination. To clarify this point, let us again examine the IC model.

Figure 3 shows a box diagram of the two-stage IC model (Di Luca et al., 2007; Domini et al., 2006; Tassinari et al., 2008). In the first stage, the disparity and velocity signals are combined into a composite r_i score, which provides the best estimate of local affine structure. In a second stage, a metric depth interpretation is assigned to r_i:$\hat{z}_i = f_\rho(r_i)$, where f_ρ is a monotonically increasing function. Bear in mind that, in general, this metric interpretation is not veridical.

At least two different noise sources can affect this perceptual process. The first source of noise is the measurement error for the disparity and velocity signals. The measurement error is assumed to be additive Gaussian noise with standard deviations σ_{d_i} and σ_{v_i} for the disparity and the velocity signals, respectively. A second source of noise comes from the errors that originate from the second stage of processing (Euclidean depth interpretation). The most parsimonious assumption is that these errors ε_D can be modeled as additive Gaussian noise as well, $\varepsilon_D \sim \mathcal{N}(0, \sigma_D)$. We will now demonstrate that, because of these two sources of noise, the just noticeable disparity and velocity increments are biased estimators of the standard deviations σ_{d_i} and σ_{v_i} of the disparity and velocity noise, respectively.

The JND is a biased estimator of disparity noise

For the sake of simplicity, assume a small range of variation for the disparity and velocity signals; also assume constant measurement noise (i.e., $\sigma_{d_i} = \sigma_d$ and $\sigma_{v_i} = \sigma_v$). These assumptions are essentially equivalent to restricting the scope of the IC model to a local analysis of a smooth surface. Consider the measurement of the disparity value d_P produced by a point P on the surface of an object. Since this measurement is subject to noise,
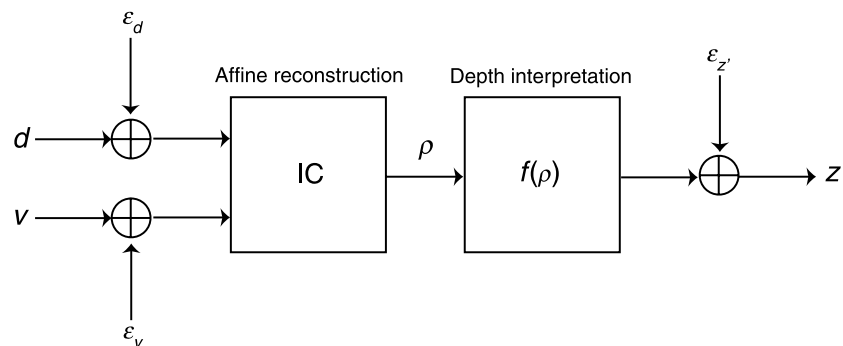

Figure 3. Box diagram of the IC model. Two sources of noise are specified: (i) the measurement noise for the disparity (ε_d) and velocity (ε_v) signals and (ii) the decision noise.

we can write $d_P = \mathbb{E}(d_P) + \varepsilon_{d_P}$. By dividing the previous equation by σ_d, we obtain

$$\frac{d_P}{\sigma_d} = \frac{\mathbb{E}(d_P)}{\sigma_d} + \frac{\varepsilon_{d_P}}{\sigma_d}. \tag{7}$$

Equation 7 can thus be rewritten as

$$r_P = \rho_P + \varepsilon_1, \tag{8}$$

where $\rho_P = \mathbb{E}(r_P)$ and $\varepsilon_1 \sim \mathcal{N}(0, 1)$.

The metric interpretation stage of the IC model can be described through a function $f_\rho(r_P)$, which imposes a metric interpretation on r_P. We thus assume that the perceived depth $\hat{z}_P$ is corrupted by a further source of Gaussian noise, $\varepsilon_D \sim \mathcal{N}(0, \sigma_D)$:

$$\hat{z}_P = f_\rho(r_P) + \varepsilon_D. \tag{9}$$

By performing a first-order Taylor expansion of the function $f_\rho(r)$ around ρ_P, we obtain

$$f_\rho(r_P) \approx f_\rho(\rho_P) + f_{\rho'}(\rho_P)(r_p - \rho_P). \tag{10}$$

By using Equation 9, we can write

$$\hat{z}_P \approx f_\rho(\rho_P) + f_{\rho'}(\rho_P)(r - \rho_P) + \varepsilon_D. \tag{11}$$

Since $\varepsilon_1 = r_P - \rho_P$ (see Equation 8) represents the perturbation around ρ_P, Equation 11 becomes

$$\hat{z}_P \approx f_\rho(\rho_P) + f_{\rho'}(\rho_P)\varepsilon_1 + \varepsilon_D. \tag{12}$$

In Equation 12, the first term represents the expected value of $\hat{z}_P$, whereas the other two terms are independent random variables having zero mean and standard deviations $f'_\rho(\rho_P)$ (since ε_1 has unit variance) and σ_D, respectively. Therefore

$$\sigma_{\hat{z}} = \sqrt{\left[f'_\rho(\rho_P)\right]^2 + \sigma_D^2}. \tag{13}$$

Equation 13 indicates that depth discrimination depends both on the variance σ_D^2 of the depth-interpretation noise and on the slope of the function $f_\rho(\rho)$ at ρ_P.

To allow an above-threshold depth discrimination, perceived depth must be increased by an amount $\Delta\hat{z} = \sigma_{\hat{z}}$. An increment equal to $\Delta\hat{z}$ is produced by increasing ρ_P by some amount $\Delta\rho$:

$$\Delta\hat{z} = f_\rho(\rho_P + \Delta\rho) - f_\rho(\rho_P). \tag{14}$$

By performing a first-order Taylor expansion of the function $f_\rho(\rho_P)$ around ρ_P, the previous equation can be approximated by

$$\Delta\hat{z} \approx f_\rho(\rho_P) + f'_\rho(\rho_P)\Delta\rho - f_\rho(\rho_P) = f'_\rho(\rho_P)\Delta\rho. \tag{15}$$

If we now equate Equations 13 and 15, we obtain

$$f'_\rho(\rho_P)\Delta\rho = \sqrt{\left[f_{\rho'}(\rho_P)\right]^2 + \sigma_D^2}. \tag{16}$$

From the previous equation, we can derive the increment $\Delta\rho$ corresponding to one JND of perceived depth:

$$\Delta\rho = \sqrt{1 + \frac{\sigma_D^2}{\left[f'_\rho(\rho_P)\right]^2}}. \tag{17}$$

Since only the disparity signal is present, $\Delta\rho = \frac{\mathbb{E}(\Delta d)}{\sigma_d}$. One JND will require an increase in the disparity signals by an amount equal to

$$\mathbb{E}(\Delta d) = \sigma_d \sqrt{1 + \frac{\sigma_D^2}{\left[f_{\rho'}(\rho_P)\right]^2}}. \tag{18}$$

In conclusion, the JND computed from stimuli providing only disparity information is a *biased estimator* of the standard deviation of the disparity noise. In fact, the disparity increment $\mathbb{E}(\Delta d)$ corresponding to one JND estimates the standard deviation of the measurement noise σ_d only up to a multiplicative bias factor equal to $\sqrt{1 + \frac{\sigma_D^2}{[f_{\rho'}(\rho_P)]^2}}$.

Since $f_\rho(\rho)$ could be any non-linear function, it is important to note that the JND can vary with the intensity of the disparity signals d_i, *even if the measurement noise of the disparity signals (ε_{d_i}) and the noise due to the depth interpretation (ε_D) are kept constant.* If $f_\rho(\rho)$ were a decelerating function, for example, the JND would be an increasing function of d_i.

According to the IC model, therefore, the JND is not informative about the standard deviation of the measurement noise. The fact that the JND varies with the intensity of the disparity signals does not necessarily mean that the same relation holds for the standard deviation of the measurement noise. It follows that any test of the IC model that posits $\mathrm{E(JND)} = \sigma_d$ must be taken with a grain of salt.

The results of Farell, Li, and Mckee (2004a, 2004b), who employed a disparity range similar to that of our stereo displays (0–10′), are also relevant to the present discussion. When standard and test stimuli were presented in successive intervals of a 2-IFC task, they found that disparity-discrimination thresholds increased with the pedestal disparity. When standard and test stimuli were embedded within the same display, however, they found that discrimination thresholds remained constant. This discrepancy suggests that a two-alternative forced-choice discrimination task may introduce an additional source of noise, since the 3D structure perceived in the second

interval must be compared to a stored memory representation of the 3D structure perceived in the first interval.

A proper test of the IC model

The previous considerations allow us to formulate a proper test for the IC model. Consider two stimuli, one providing disparity-only information and one providing velocity-only information. If the velocity-only stimulus $z_v^{(j)}$ is perceptually matched in depth to the disparity-only stimulus $z_d^{(j)}$, then, according to the IC model, the SNRs of the two stimuli (see Equation 3) must be equal.

To compute the SNR for the two stimuli, we reason as follows. The signal intensities $v^{(j)}$ and $d^{(j)}$ are provided by the stimulus displays. In The JND is a biased estimator of disparity noise section, however, we have shown that the JND is a biased estimator of the standard deviation of the measurement noise. Nevertheless, this is not a problem for the present purposes, if we assume the same $f_\rho(\rho)$ function in the bias factor $\sqrt{1+\frac{\sigma_D^2}{[f_\rho'(\rho_P)]^2}}$ for both the disparity and velocity signals. To test the IC model, therefore, it is sufficient to check whether

$$\frac{d^{(j)}}{\text{JND}_{d^{(j)}}} = \frac{v^{(j)}}{\text{JND}_{v^{(j)}}}, \qquad (19)$$

where $\text{JND}_{d^{(j)}}$ and $\text{JND}_{v^{(j)}}$ are the JNDs estimated at the signals' intensities $d^{(j)}$ and $v^{(j)}$, respectively.

Equation 19 can also be expressed in terms of simulated depth magnitudes, rather than in terms of disparity and velocity signals. First, notice that $v^{(j)} = \omega z_v^{(j)}$ and $d^{(j)} = \mu z_d^{(j)}$. Second, notice that $\text{JND}_{v^{(j)}} = \omega \text{JND}_{z_v^{(j)}}$ and $\text{JND}_{d^{(j)}} = \mu \text{JND}_{z_d^{(j)}}$. Remember that by $\text{JND}_{d^{(j)}}$ and $\text{JND}_{v^{(j)}}$ we mean the estimated discrimination threshold expressed in terms of the signal intensities d and v; by $\text{JND}_{z_d^{(j)}}$ and $\text{JND}_{z_v^{(j)}}$ we mean the same discrimination thresholds, but this time as expressed in terms of the simulated depth magnitudes. Equation 19, therefore, can be written as

$$\frac{\mu z_d^{(j)}}{\mu \text{JND}_{z_d^{(j)}}} = \frac{\omega z_v^{(j)}}{\omega \text{JND}_{z_v^{(j)}}}. \qquad (20)$$

In conclusion, the predictions of the IC model concerning the issue of Fechnerian scaling can be tested by checking whether

$$z_d^{(j)} = z_v^{(j)} \frac{\text{JND}_{z_d^{(j)}}}{\text{JND}_{z_v^{(j)}}}. \qquad (21)$$

IC and Fechnerian scaling

Now, let us consider the issue of whether the measurement noise is constant, or whether it varies with signal intensity. Suppose that a disparity-only cylinder is perceptually matched to a motion-only cylinder. Note that, for the two stimuli to be perceived with the same depth elongation, it is not sufficient that they are both matched in simulated depth. Perceived depth from single-cues or combined-cues stimuli can, in fact, be under-estimated or overestimated, depending on the viewing parameters of the visual scene (that is, the fixation distance, for disparity information, and the 3D angular velocity, for motion information).

Let $d^{(1)}$ be the front-to-back disparity of the two cylinders and let $v^{(1)}$ be the front-to-back relative velocity. According to the IC model, it must be true that $\rho_d^{(1)} = \rho_v^{(1)}$, or equivalently that $\frac{d^{(1)}}{\sigma_{d^{(1)}}} = \frac{v^{(1)}}{\sigma_{v^{(1)}}}$. Now, suppose that the disparity signal is increased by one standard deviation $\sigma_{d^{(1)}}$ of measurement noise. The ensuing value $\rho_{d^{(2)}}$ is

$$\rho_d^{(2)} = \frac{d^{(1)} + \sigma_{d^{(1)}}}{\sigma_{d^{(2)}}} = \frac{\frac{d^{(1)}}{\sigma_{d^{(1)}}} + \frac{\sigma_{d^{(1)}}}{\sigma_{d^{(1)}}}}{\frac{\sigma_{d^{(2)}}}{\sigma_{d^{(1)}}}} = \left[\rho_d^{(1)} + 1\right]\frac{\sigma_{d^{(1)}}}{\sigma_{d^{(2)}}}. \qquad (22)$$

The measurement noise σ_d may remain constant, or it may vary with signal intensity. In the following, we will consider the consequences that would follow each of these two possibilities:

1. If the noise of the disparity measurement is constant, then $\sigma_{d^{(1)}} = \sigma_{d^{(2)}}$, $\rho_d^{(2)} = \rho_d^{(1)} + 1$, and $\rho_v^{(2)} = \rho_v^{(1)} + 1$. The increase of one JND produces a unit increase of ρ for both stereo-only and motion-only displays. After increasing the motion and stereo signals by one JND, therefore, the two stimuli should still be perceived as having the same depth elongation. If the stimulus pair $\{z_d^{(m)}, z_v^{(m)}\}$ is perceptually matched in depth, and so is the stimulus pair $\{z_d^{(n)}, z_v^{(n)}\}$, then the stimuli $z_d^{(m)}$ and $z_d^{(n)}$ should be separated by the same number of JNDs in depth as the stimuli $z_v^{(m)}$ and $z_v^{(n)}$. We want to stress that this prediction, which MacKenzie et al. (2008) attribute to the IC model, holds only if the measurement noise for the disparity and velocity signals does not vary with signal intensity. In their own data, MacKenzie et al. (2008) found that the JNDs for the velocity and disparity stimuli do indeed vary as a function of signal intensity. Any conclusion based on the assumption of a constant measurement noise, therefore, is questionable.

2. Now, consider the case in which the measurement noise varies with signal intensity. Would the IC model be falsified in this case? Not at all, *unless the standard deviation of the measurement noise varied at the same rate as the signal intensity*. In this case, the ratio between the signal intensity and the standard deviation of the measurement noise would remain constant, and according to the IC model, all stimuli would be perceived as having the same

depth extent. The above consideration has an important implication: according to the IC model, the Weber law does not apply within the range used in the psychophysical experiments on 3D depth perception (see Stairway to depth perception section).

Even though we never made the assumption of constant noise, the paper on which the authors based their investigation may have been ambiguous in this respect (see Domini et al., 2006). Even if it had been explicitly made, however, the assumption of constant noise would not be critical; it would simply limit the scope of the IC model to a local analysis.

An empirical test of the IC model

We tested Fechner's hypothesis in the domain of depth perception by following the procedure depicted by the right panel of Figure 1, that is, by adding successive JNDs in a step-by-step fashion. Our stimulus was composed of three dotted vertical lines embedded in a cloud of random dots (see Figure 4). Two *flanking* lines were positioned at fixation distance. A third line, which projected midway between the two, was located in depth in front of the flankers. Participants were asked to judge the depth separation (which we will call *stimulus depth*) between the flankers and the central line. The 3D information was provided either by binocular disparities (*stereo stimulus*) or by image velocities (*motion stimulus*).

At the beginning of the experiment, the depth $z_v^{(1)}$ of the motion stimulus was set at 12.5 mm. For each participant, through a staircase procedure, we found the simulated depth of a stereo-only stimulus, $z_d^{(1)}$, which was perceptually matched in depth to the motion-only stimulus.

Having found the two "starting points" $z_v^{(1)}$ and $z_d^{(1)}$, we built two psychophysical scales, one for the motion-only stimuli and one for the stereo-only stimuli. The discrimination thresholds measured at $z_v^{(1)}$ and $z_d^{(1)}$, denoted with $\text{JND}_{z_v^{(1)}}$ and $\text{JND}_{z_d^{(1)}}$, provided the first step for the two psychophysical scales. Two new discrimination thresholds were then estimated at $z_v^{(2)} = z_v^{(1)} + \text{JND}_{z_v^{(1)}}$ and $z_d^{(2)} = z_d^{(1)} + \text{JND}_{z_d^{(1)}}$. By increasing the simulated depth by one JND at a time, we proceeded by six steps, so as to find the $z_v^{(j)}$ and $\text{JND}_{z_v^{(j)}}$ magnitudes of the motion scale, and the $z_d^{(j)}$ and $\text{JND}_{z_d^{(j)}}$ magnitudes of the stereo scale, with $j = 1,\ldots,6$. With the exception of the first step, the two sequences of $z^{(j)}$ and $\text{JND}_{z^{(j)}}$ were independently estimated for the stereo-only and motion-only stimuli.

To test the prediction of classical Fechnerian theories, we asked whether corresponding steps of the two scales evoked the same magnitudes of perceived depth. A staircase procedure was run at each step of the motion-only psychophysical scale. By means of a 2-IFC task, participants were asked to match the perceived depth of a fixed motion-only stimulus with the perceived depth of a varying stereo-only stimulus. In this manner, at each step of the motion-only psychophysical scale, we found the PSE of a perceptually matched stereo-only stimulus. We expected that the PSEs found in this manner would be equal to the $z_d^{(j)}$ values comprising the disparity-only scale.

Experiment

Methods

Observers

Four observers with normal or corrected-to-normal vision participated in the experiment. Three observers were naive to the purpose of the experiment (two graduate
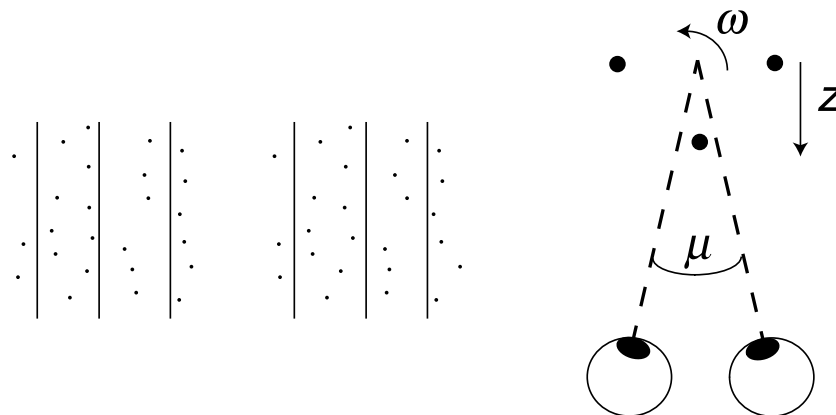

Figure 4. (Left panel) Stereogram representing a simplified version of the stimulus used in the experiment (cross-fuse). (Right panel) Schematic representation of the viewing geometry of the stimulus used in the experiment. The three dots represent a bird's-eye view of the three vertical lines shown in each stimulus display. In the figure, the central line is closer to the observer than the flanking lines. The depth separation between the central and the flanking lines is denoted by z. ω represents the angular rotation about the fixation point and μ is the vergence angle.

and one undergraduate Brown University students), and one was the first author.

Apparatus

Stereoscopic stimuli were displayed on a haploscope consisting of two CRT monitors (0.22-mm dot pitch) located on swing arms pivoting directly beneath the observer's eyes. Anti-aliasing and spatial calibrating procedures allowed spatial precision of dot location greater than hyperacuity levels. Each monitor was seen in a mirror by one eye. Head position was fixed with a chin-and-forehead locating apparatus. The actual distance from each eye to the corresponding monitor was 95 cm. The eyes' vergence was directly manipulated by physically moving the monitors on their swing arms. Since the monitors and mirrors pivot rigidly about the eye's axis of rotation, the retinal images always remain the same for all positions of the two CRT monitors. Thus, changes in eye position were dissociated from changes in retinal images.

Stimuli

The stimuli were 800 high-luminance anti-aliased dots displayed against a low-luminance background. Four hundred dots (scattered in 2D projection) were superimposed on three invisible vertical lines, each 50 mm long. To help stereoscopic fusion, the other points were randomly positioned within a volume 50 mm wide, 50 mm high, and 25 mm deep. Figure 4 provides a cartoon example of the display. One of the three vertical lines was positioned at the center of the stimulus display. The other two vertical lines were positioned at 12.5 mm to either side of the central line. The overall stimulus subtended about 2.9° of visual angle.

Depth information was provided by either disparity or velocity cues. Disparities were calculated so as to simulate a 3D structure viewed at 100 cm from the observer. The vergence angle was computed for each observer by taking into account her or his inter-ocular distance. The 2D motion of the dots in the display was computed by simulating a rotation of the simulated 3D structure about a horizontal axis positioned at fixation. The 3D structure rotated back and forth by 14°. The duration of an entire oscillation cycle was 2 s. The stimulus remained on the screen until the participant terminated the trial with a key press. This time could comprise many oscillation cycles of the simulated 3D structure. For both stereo and motion stimuli, the simulated depth position of the two flanking lines was at fixation. The simulated relative depth of the central line with respect to the flanking lines was either positive (central line in front) or negative.

Procedure

Observers judged the depth separation between the two flanking lines and the central line. In a 2-IFC task,

observers were asked to determine which of the two successively presented stimuli evoked a larger depth separation. The simulated depth separation of the stimulus in one of the two intervals was constant (*comparison stimulus*), whereas the depth separation of the other stimulus was varied according to a staircase procedure (*test stimulus*). We used staircases to control the value of simulated depth and four reversal rules—3 down/1 up, 1 down/3 up, 2 down/1 up, and 1 down/2 up—to sample points along the entire psychometric function. Four staircases were used for each psychometric function, which corresponds to approximately 200 trials per function (each staircase was terminated after 6 reversals). For each observer, the JND and PSE were estimated from the fitted psychometric function. The mean and standard deviation of a cumulative normal were used to estimate PSEs and JNDs, respectively. Psychometric functions were fitted using psignifit version 2.5.6 (see http://bootstrap-software.org/psignifit/), a software package that implements the maximum-likelihood method described by Wichmann and Hill (2001).

In the first part of the experiment (*depth matching*), the *comparison* was a motion stimulus simulating a depth of 12.5 mm and the *test* was a stereo stimulus that was varied according to a staircase procedure. The purpose of this part of the experiment was to determine the PSE of the stereo stimulus perceptually matched to the motion stimulus simulating a depth of 12.5 mm. The simulated depth of 12.5 mm for the motion stimulus and the simulated depth at the PSE for the stereo stimulus were then used as the starting points for building the motion-based and stereo-based psychophysical scales.

In the second part of the experiment (*depth discrimination*), we built the motion-based and stereo-based psychophysical scales. The motion-based scale was generated by adding successive JNDs to the starting point (i.e., the simulated depth of 12.5 mm). The JNDs were estimated by a depth-discrimination task. In a 2-IFC task, observers were asked to determine which of the two successively presented stimuli appeared to be deeper. To estimate the first JND, we used a motion stimulus simulating a depth of 12.5 mm as the *comparison* and a motion stimulus that was varied according to a staircase procedure as the *test*. From the resulting psychophysical function, we estimated the first JND. In the successive step, the motion *comparison stimulus* simulated a depth of 12.5 mm plus the JND estimated in the previous step; the motion *test stimulus* was again varied according to a staircase procedure. In this way, a second JND was estimated, and this procedure was repeated five times.

The stereo-based psychophysical scale was generated in a similar manner. We started from the simulated depth magnitudes determined by the PSE obtained in the preliminary part of the experiment (*depth matching*). In this case, both the *comparison* and *test* stimuli were defined by disparity information. By using the same procedure as for the motion stimuli, we determined the

sequence of the five JNDs, which comprise the stereo-based psychophysical scale.

In the third part of the experiment (*depth matching*), observers were asked to compare motion and stereo stimuli. A 2-IFC task was performed in five blocks of trials. In each block, the *comparison* was a motion stimulus defined by one of the five simulated depth magnitudes that comprise the motion-based psychophysical scale. The *test stimulus* was a static stereo display, which was varied according to a staircase procedure. The third part of the experiment allowed us to determine the stereo-depths that perceptually matched each step of the motion-based psychophysical scale.

Results

The left panel of Figure 5 shows the psychophysical scales obtained for the motion (green) and stereo (red) stimuli. The first fact to highlight is that motion stimuli simulating a depth of 12.5 mm were perceived as deep as stereo stimuli simulating a depth of 5.3 mm, on average. This sizable mismatch between the simulated depths of the two stimuli is not surprising. A vast literature on perceived Structure from Motion (SfM) indicates that the visual system relies only on a first-order temporal analysis of the optic flow (e.g., Domini & Caudek, 1999; Caudek, Domini, & Di Luca, 2002; Caudek & Proffitt, 1993; Di Luca, Domini, & Caudek, 2004; Domini & Caudek, 2003a; Domini & Caudek, 2003b; Domini, Caudek, & Skirko, 2003; Domini, Vuong, & Caudek, 2002; Liter,

Braunstein, & Hoffman, 1993; Norman & Todd, 1993; Todd, 1998; Todd & Bressan, 1990). As a consequence, a veridical recovery of 3D Euclidean structure from motion is virtually impossible. In a series of studies, we have shown the existence of SfM metamers: very different magnitudes of depth can be made perceptually indistinguishable, when coupled with appropriate magnitudes of simulated 3D angular rotation (Domini & Caudek, 2003b; Domini, Caudek, & Proffitt, 1997). In the present experiment, in order to produce a perceptual match, we purposely chose a magnitude of 3D rotation for the motion stimuli that requires a large discrepancy between the simulated depth magnitudes of the stereo and motion displays.

The second fact to highlight in Figure 5 is that the discrepancy between the simulated depth magnitudes of the two stimuli increases at each step of their psychophysical scales. The last step of the motion scale corresponds to a simulated depth of 43 mm; the last step of the stereo scale corresponds to a simulated depth of 15 mm. The discrimination threshold increases with simulated depth at a faster rate within the motion scale than within the stereo scale (see Figure 5, left panel).

The five depth magnitudes that were simulated for the motion stimuli were equal to the depth magnitudes defining the five steps of the motion-based psychophysical scale, as determined in the second part of the experiment (Figure 6, left panel, green line). For each of these five motion-defined depths, we estimated the PSE for the stereo stimuli. The PSEs of the stereo stimuli are marked in Figure 6 as green squares. Note how similar these values are to those obtained by independently building the stereo scale (Figure 6, red line, left and right panels).
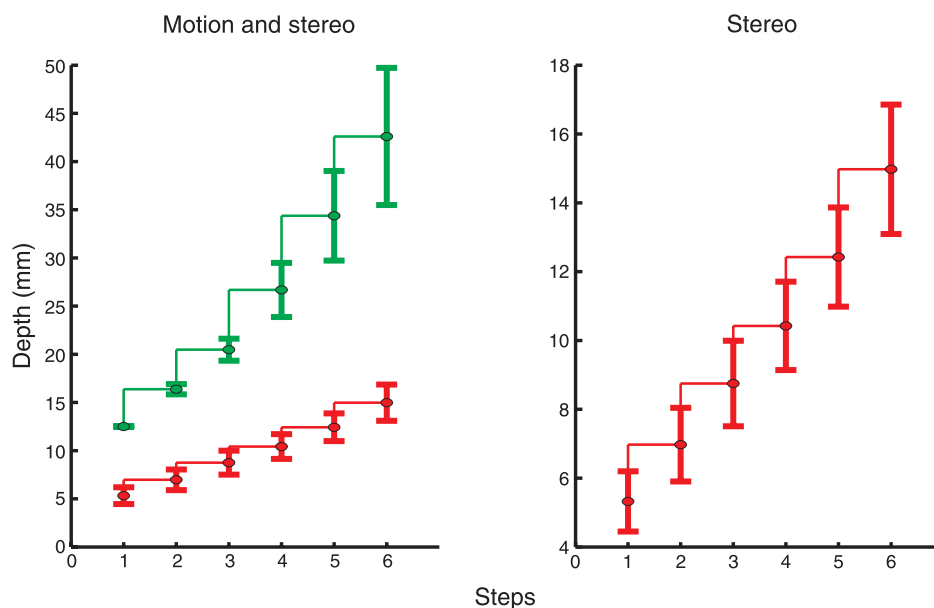

Figure 5. The left panel shows the psychophysical scales constructed from the motion (green lines) and stereo (red lines) cues. Each scale corresponds to the cumulation of psychometric increments (JNDs) measured in the depth-discrimination part of the experiment. The reported values are averaged over four observers. Vertical bars indicate ± one standard error of the mean. (Right panel) Enlarged version of the psychophysical scale derived from the stereo-depth increments.
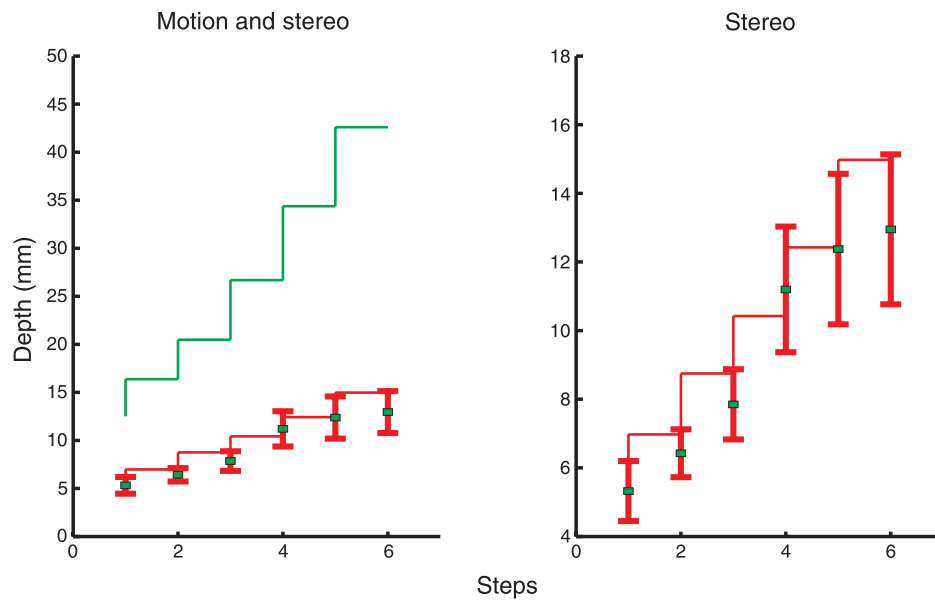
Figure 6. In the third part of the experiment, observers were asked to compare motion and stereo stimuli. The left panel represents the magnitudes of stereo-depth (green squares), which were required to be perceptually matched to the simulated depth magnitudes defined by the successive psychometric increments (JNDs) of the motion-based psychophysical scale. Vertical bars indicate ± one standard error of the mean. The solid lines represent the psychophysical scales constructed in the second part of the experiment (green: motion; red: stereo). (Right panel) Enlarged version of the stereo-depth matches.

Having said that, the critical question is whether these two sequences of largely discrepant simulated depth magnitudes give rise to *the same amount of perceived depth* at each step of the psychophysical scales. The answer to this question is provided by Figure 6, where we have plotted the results of the third part of the experiment. Remember that, in part three of the experiment, observers

compared static stereo stimuli with motion stimuli. In a 2-IFC, the motion stimulus was kept fixed, whereas the stereo stimulus was varied according to a staircase procedure.

The mean difference between the simulated depth magnitudes of the PSEs for the stereo stimuli computed in part 2 and part 3 of the experiment was not significant:
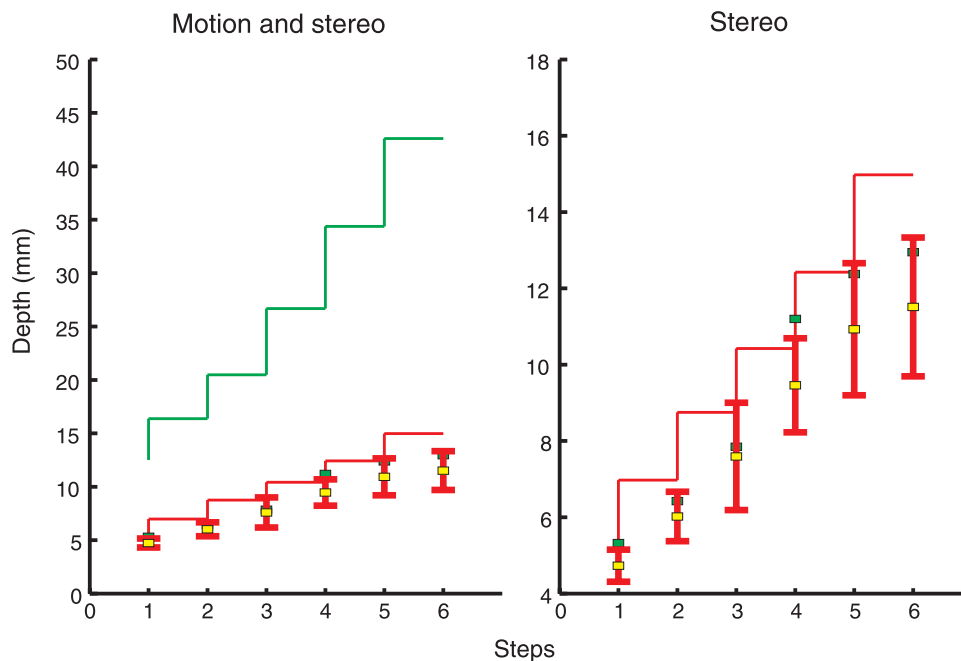

Figure 7. Data in Figure 6 replotted together with the predictions of the IC model (Equation 21, yellow squares). Vertical bars indicate ± one standard error.

$\bar{Y}_2 - \bar{Y}_3 = 1.0743$ mm (95% C.I.: -0.4242, 1.3409; $t_{23} = 1.0743$, $p > 0.05$). We can thus conclude that JND increments applied independently to stereo or motion stimuli correspond to equivalent increments in perceived depth. These results are therefore compatible with the Fechnerian theory relating perceived magnitudes to JND sums.

The data in Figure 6 have been replotted in Figure 7 together with the predictions of the IC model (see Equation 21). Note the good agreement between the theoretical predictions (which do not involve any free parameters) and the experimental data. The predictions of the IC model are, in fact, formulated only in terms of the PSEs and the JNDs estimated in the experiment.

A statistical test was provided by a linear regression on the PSEs found in part three of the experiment. The predictor was computed, as indicated in Equation 21, by the simulated depth magnitudes $z_v^{(j)}$ of the motion stimuli of part two of the experiment, which were weighted by the ratio $\frac{\text{JND}_{z_d^{(j)}}}{\text{JND}_{z_v^{(j)}}}$, with $j = 1,\ldots,6$. If the IC model is correct, then we expect a linear relationship with zero intercept and a slope of one. The IC model can be contrasted with a model assuming an unbiased derivation of 3D Euclidean shape from retinal cues. According to such a model, $\text{PSE}_{z_d^{(j)}} = \text{PSE}_{z_v^{(j)}}$. To take the individual differences into account, we centered the data: the mean was subtracted from the data of each subject, both for the response variable $z_d^{(j)}$ and for the predictors of the two models. The centered data were then analyzed by linear regression.

By using the predictor defined by the IC model, the slope for the regression model was equal to 0.9185 (95% C.I.: 0.5726, 1.2644). By using the predictor of a model assuming an unbiased derivation of 3D Euclidean shape from retinal cues, the slope of the regression line was equal to 0.2646 (95% C.I.: 0.2026, 0.3266). It is clear that the IC model, which posits a link between perceived depth and depth discrimination, is capable of a better prediction of the data. We can thus conclude by saying that these results support the claim of the IC model of a strong link between perceived depth and depth discrimination.

General discussion

MacKenzie et al. (2008) proposed an indirect way of testing the IC model by pointing out that, according to our model, the JND provides an adequate metric for perceived depth. In their experiment, MacKenzie et al. found that perceived depth and depth discriminability vary independently, contrary to what is predicted by the IC model. MacKenzie et al. maintained that "[t]ests of the relationship between JND counts, sensory magnitudes, and stimulus intensity along other perceptual dimensions (e.g., brightness and loudness) have found that a simple sum of JND's does not predict the resulting change in

sensory magnitudes (e.g., Newman, 1933; Stevens, 1957; Stevens, 1961). Our results are consistent with this literature". MacKenzie et al., therefore, concluded that "IC is inconsistent with the psychophysics of depth perception."

With the present empirical investigation, by using a different JND counting procedure, we demonstrate that the JNDs do indeed provide a unit of measurement for depth perception: within the range of the present stimulus settings, the separation between two objects as measured in JNDs does predict their separation in perceived depth.

MacKenzie et al. accompanied their empirical research with a theoretical discussion concerning the Modified Weak Fusion (MWF) theory of depth-cue integration (Landy, Maloney, Johnston, & Young, 1995). They argued that the IC model shares with the MWF model the same desirable characteristic of "combin[ing] two cues with different means and standard deviations [through a weighted sum] in order to maximize the Signal-to-Noise Ratio (SNR) of the resulting decision variable." This statement, however, is incomplete, because the combination rules proposed by the IC and MWF models are *not equivalent* in three respects.

First, the combination rule proposed by the MWF model *does not,* in general, maximize the SNR of the combined estimate; this happens only if the estimates derived from the single depth cues are unbiased (see Appendix A).

Second, the weighted average computed by the IC model, in contrast to the MWF, does not require the estimation of viewing parameters not specified by optical information. In the IC model, in fact, a depth interpretation is not provided for each cue in isolation. Therefore, the estimate of parameters such as the vergence angle, for example, is not necessary. Instead, according to IC, the image signals are first combined in the composite score ρ. Then, a maximum likelihood depth interpretation is provided for ρ.

Finally, the IC model predicts a strong link between perceived depth and depth discrimination, whereas MWF does not. For the MWF, perceived depth is an unbiased estimate of the "true" depth (Oruç, Maloney, & Landy, 2003). No attempt has ever been made by MWF proponents to model the systematic misperceptions of perceived depth magnitudes in terms of the discrimination thresholds.

Conclusions

The IC model predicts a strong link between perceived depth and depth discrimination, namely that the perceived depth separation between any two single-cue stimuli should depend on the number of JNDs separating the two stimuli. The present investigation provides support for this hypothesis.

Appendix A

In the Local affine structure and signal-to-noise ratio section, we showed that the precision of an affine estimate can be expressed in terms of the SNR. Since the goal of the first stage of the IC model is to estimate the local affine structure, it is important to evaluate the performance of the MWF model in this respect. To do so, we will present cases in which the single-cue stimuli provide unbiased or biased estimates of the 3D shape. In both cases, we will ask whether the MWF provides a "statistically optimal" combination of cues. The answer to this question is that the MWF combination rule is optimal (in terms of SNR maximization) only if the single-cue estimates are unbiased.

Unbiased estimation from single cues

Assume that the single-cue estimates are unbiased: $\mathbb{E}(\hat{z}_{d_i}) = \mathbb{E}(\hat{z}_{v_i}) = z_i$. Since the weights of the MWF combination rule sum to 1, the expected value of the combined estimate $(\hat{z}_{c_i})$ will also be unbiased: $\mathbb{E}(\hat{z}_{c_i}) = \mathbb{E}(\hat{z}_{d_i}) = \mathbb{E}(\hat{z}_{v_i}) = z_i$. The combination rule of the MWF model minimizes the variance $\sigma^2_{\hat{z}_{c_i}}$ of the combined estimate $\hat{z}_{c_i}$ and, therefore, also maximizes the SNR of the final estimate, $\frac{\mathbb{E}(\hat{z}_{c_i})}{\sigma_{\hat{z}_{c_i}}}$. In fact, the expected value $\mathbb{E}(\hat{z}_{c_i})$ of the combined estimate does not depend on the choice of the weights.

Unbiased estimates of the depth map z_i from stereo information are possible only if the vergence angle μ can be accurately estimated (see Equation 1). Correspondingly, a veridical interpretation of motion information requires an unbiased estimate of the 3D angular rotation ω (see Equation 2). The psychophysical literature, however, has revealed the existence of large and systematic biases in the perceptual interpretation of 3D shape from binocular disparities and retinal velocities. In a series of papers, we showed that the visual system is unable to recover unbiased estimates of 3D rotation from the velocity field (Caudek & Domini, 1998; Caudek & Rubin, 2001; Domini, Caudek, & Richman, 1998; Domini, Caudek, Turner, & Favretto, 1998). Regarding disparity information, similar results have also been found for the vergence angle (Johnston, 1991; Johnston, Cumming, & Landy, 1994).

Proponents of MWF agree that it is problematic to derive unbiased estimates of the viewing parameters from single cue in isolation. In the original version of the MWF model, they argued that unbiased estimates of μ and ω may be obtained by *promotion,* that is, by relying on the mutual constraints deriving from the simultaneous presence of disparity and velocity information (see Richards & Lieberman, 1985). If a process akin to promotion were to take place, observers' accuracy would improve when more depth cues are added to the stimulus displays. In our own research, however, we have shown that the exact opposite can happen Tassinari et al. (2008). In sum, neither the estimation of the viewing parameters nor the process of promotion guarantees unbiased perceptual estimates of 3D shape.

Biased estimation from single cues

Assume that the single-cue estimates are biased. For the disparity and velocity cues, we can then write

$$\hat{z}_{d_i} = \xi_{d_i} + \varepsilon_{z_{d_i}}, \tag{A1}$$

$$\hat{z}_{v_i} = \xi_{v_i} + \varepsilon_{z_{v_i}}, \tag{A2}$$

where $\xi_{d_i} = z_i + \text{bias}_d$ and $\xi_{v_i} = z_i + \text{bias}_v$. MacKenzie et al. showed that, in order to maximize the SNR, the weights of the combination rule must be proportional to the means and inversely proportional to the variances of the individual estimates: $w_{\hat{z}_d} \propto \frac{\xi_{d_i}}{\sigma^2_{z_{d_i}}}$ and $w_{\hat{z}_v} \propto \frac{\xi_{v_i}}{\sigma^2_{z_{v_i}}}$. Note, however, that the weights of the MWF combination rule depend only on the variances of the depth estimates: $w_{\hat{z}_d} \propto \frac{1}{\sigma^2_{z_{d_i}}}$ and $w_{\hat{z}_v} \propto \frac{1}{\sigma^2_{z_{v_i}}}$. It then follows that, when the individual estimates are biased, the MWF combination rule is not optimal: the combined estimate does not maximize the SNR. To provide an intuitive example, Figure A1 shows the results of a simulation in which two biased single-cue depth estimates are combined according to the MWF combination rule. In this example, a greater magnitude of depth is estimated from stereo-only than from motion-only. The stereo estimate, however, has a larger standard deviation than the motion estimate. The MWF combination rule, therefore, weights motion more heavily than stereo; the estimate from the stereo-motion cues (yellow) is more similar to the estimate that is derived from the motion-only cue than to the estimate that is derived from the stereo-only cue.

The green and red colors code the depth estimates derived from the disparity-only and velocity-only cues, respectively. The combined-cues depth estimate is represented in yellow and is computed according to the MWF combination rule. The solid black lines represent the expected values of the depth estimates. The green, red, and yellow bands represent the standard deviation of the depth estimates: each band represents the expected value of the depth estimate ± one standard deviation. The depth magnitudes derived from the stereo cue are larger than those derived from the velocity cue. The standard deviation of the depth estimate derived from stereo, however, is larger than the standard deviation of the depth estimate derived from motion. According to MWF, therefore, the motion cue is weighted more heavily than the
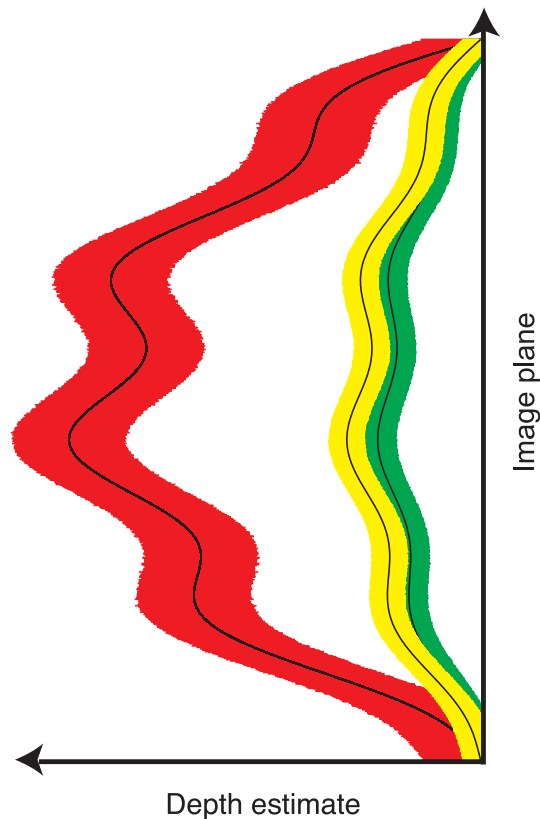
Figure A1. Results of a simulation in which two biased single-cue depth estimates are combined by the MWF combination rule.

stereo cue. Note that: (i) the width of the yellow band is smaller than each of the other two, and (ii) the combined-cues depth estimate is closer to the depth estimate derived from the motion cue.

The difference between the two models is that the IC model weights the signal with larger SNR more heavily, whereas the MWF model weights the signal with smaller noise variance more heavily. As a consequence, the SNR of the combined-cues estimate will be larger for the IC model.

In conclusion, MWF is a less precise estimator of affine structure than IC. The only exception is when unbiased estimates can be derived from single cues. In those circumstances, the two combination rules are equivalent. However, such cases are very rare, as suggested by previous empirical evidence and by the results of the present investigation.

References

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 609–621. [PubMed]

Caudek, C., Domini, F., & Di Luca, M. (2002). Short-term temporal recruitment in structure from motion. *Vision Research, 42,* 1213–1233. [PubMed]

Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 32–47. [PubMed]

Caudek, C., & Rubin, N. (2001). Segmentation in structure from motion: Modeling and psychophysics. *Vision Research, 41,* 2715–2732. [PubMed]

Di Luca, M., Domini, F., & Caudek, C. (2004). Spatial integration in structure from motion. *Vision Research, 44,* 3001–3013. [PubMed]

Di Luca, M., Domini, F., & Caudek, C. (2007). The relation between disparity and velocity signals of rigidly moving objects constraints depth order perception. *Vision Research, 47,* 1335–1349. [PubMed]

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 426–444. [PubMed]

Domini, F., & Caudek, C. (2003a). Perception of slant and angular velocity from a linear velocity field: Modeling and psychophysics. *Vision Research, 43,* 1753–1764. [PubMed]

Domini, F., & Caudek, C. (2003b). 3-D structure perceived from dynamic information: A new theory. *Trends in Cognitive Sciences, 7,* 444–449. [PubMed]

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1111–1129. [PubMed]

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60,* 1164–1174. [PubMed]

Domini, F., Caudek, C., & Skirko, P. (2003). Temporal integration of motion and stereo cues to depth. *Perception & Psychophysics, 65,* 48–57. [PubMed]

Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research, 46,* 1707–1723. [PubMed]

Domini, F., Caudek, C., Turner, J. & Favretto, A. (1998). Discriminating constant from variable angular velocities in structure from motion. *Perception & Psychophysics, 60,* 747–760. [PubMed]

Domini, F., Vuong, Q. C., & Caudek, C. (2002). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 816–838. [PubMed]

Enright, J. T. (1991). Stereo-thresholds: Simultaneity, target proximity and eye movements. *Vision Research, 31,* 2093–2100. [PubMed]

Farell, B., Li, S., & McKee, S. P. (2004a). Coarse scales, fine scales and their interactions in stereo vision. *Journal of Vision, 4*(6):8, 488–499, http://journalofvision.org/4/6/8/, doi:10.1167/4.6.8. [PubMed] [Article]

Farell, B., Li, S., & McKee, S. P. (2004b). Disparity increment thresholds for gratings. *Journal of Vision, 4*(3):3, 156–168, http://journalofvision.org/4/3/3/, doi:10.1167/4.3.3. [PubMed] [Article]

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31,* 1351–1360. [PubMed]

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34,* 2259–2275. [PubMed]

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, Optics and Image Science, 8,* 377–385. [PubMed]

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception, 22,* 1441–1465. [PubMed]

MacKenzie, K. J., Murray, R. F., & Wilcox, L. M. (2008). The intrinsic constraint approach to cue combination: An empirical and theoretical evaluation. *Journal of*

Vision, 8(8):5, 1–10, http://journalofvision.org/8/8/5/, doi:10.1167/8.8.5. [PubMed] [Article]

Newman, E. B. (1933). The validity of the just noticeable difference as a unit of psychological magnitude. *Transactions of the Kansas Academy of Science, 36,* 172–175.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53,* 279–291. [PubMed]

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43,* 2451–2468. [PubMed]

Richards, W., & Lieberman, H. R. (1985). Correlation between stereo ability and the recovery of structure-from-motion. *American Journal of Optometry and Physiological Optics, 62,* 111–118. [PubMed]

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181. [PubMed]

Stevens, S. S. (1961). To honor Fechner and repeal his law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science, 133,* 80–86. [PubMed]

Tassinari, H., Domini, F., & Caudek, C. (2008). The intrinsic constraint model for stereo-motion integration. *Perception, 37,* 79–95. [PubMed]

Todd, J. T. (1998). Theoretical and biological limitations on the visual perception of 3D structure from motion. In T. Watanabe (Ed.), *High-level motion processing—Computational, neurophysiological and psychophysical perspectives* (pp. 359–380). Cambridge, MA: MIT Press.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430. [PubMed]

Todd, J. T., Oomes, A. H., Koenderink, J. J., & Kappers, A. M. (2001). On the affine structure of perceptual space. *Psychological Science, 12,* 191–196. [PubMed]

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313. [PubMed]