The intrinsic constraint approach to cue combination: An empirical and theoretical evaluation

Kevin J. MacKenzie  
Centre for Vision Research and Department of Psychology,  
York University, Toronto, ON, Canada, &  
School of Psychology, Bangor University, Bangor, UK

Richard F. Murray  
Centre for Vision Research and Department of Psychology,  
York University, Toronto, ON, Canada

Laurie M. Wilcox  
Centre for Vision Research and Department of Psychology,  
York University, Toronto, ON, Canada

We elucidate two properties of the intrinsic constraint (IC) model of depth cue combination (F. Domini, C. Caudek, & H. Tassinari, 2006). First, we show that IC combines depth cues in a weighted sum that maximizes the signal-to-noise ratio of the combined estimate. Second, we show that IC predicts that any two depth-matched pairs of stimuli are separated by equal numbers of just noticeable differences (JNDs) in depth. That is, IC posits a strong link between perceived depth and depth discrimination, much like some Fechnerian theories of sensory scaling. We test this prediction, and we find that it does not hold. We also find that depth discrimination performance approximately follows Weber’s law, whereas IC assumes that depth discrimination thresholds are independent of baseline stimulus depth.

Keywords: depth perception, cue combination, modified weak fusion, intrinsic constraint


Introduction

Many visual cues carry information about the depth of an object, and an active topic of research is how the human visual system combines two or more such cues to arrive at a single estimate of depth. Most studies of cue combination have used the framework of the modified weak fusion (MWF) model (Landy, Maloney, Johnston, & Young, 1995; Maloney & Landy, 1989), but recently Domini, Caudek, and Tassinari (2006) have proposed an alternative account of depth cue combination, which they call the intrinsic constraint (IC) model. In this article, we derive two properties of the IC model and test one of them experimentally. First, we show that IC effectively combines depth cues in an optimal weighted sum. Second, we show that IC predicts a strong link between perceived depth and depth discrimination, and we report a psychophysical experiment that does not support this prediction.

Cue combination models

MWF, the most widely used model of cue combination, assumes that the visual system has access to two or more depth cues that give metric depth estimates in a common unit of measurement. These cues are assumed to be degraded by statistically independent Gaussian noise. The cues are represented as random variables, say \( B(z) \) for depth from binocular disparity and \( M(z) \) for depth from structure-from-motion, where \( z \) is the true depth of the point of interest. MWF posits that the visual system combines these cues in a weighted sum: \( C(z) = w_B B(z) + w_M M(z) \). The weights \( w_B \) and \( w_M \) are non-negative and sum to one, but otherwise they are arbitrary constants. However, if the weights are set to \( w_B = \sigma_B^{-2}/(\sigma_B^{-2} + \sigma_M^{-2}) \) and \( w_M = \sigma_M^{-2}/(\sigma_B^{-2} + \sigma_M^{-2}) \), where \( \sigma_B = SD[B(z)] \) and \( \sigma_M = SD[M(z)] \), then the sum is optimal in the sense that it maximizes the signal-to-noise ratio (SNR) of the combined estimate \( C(z) \), defined as \( SNR[C(z)] = E[C(z)]/SD[C(z)] \). Here, \( E \) denotes expected value and \( SD \) denotes standard deviation (for an extensive review, see Landy et al., 1995).

In this very brief summary, we have highlighted the aspects of MWF that are most relevant to our discussion of IC. Important parts of MWF that we will not discuss at length include a promotion stage, in which direct retinal measurements of depth cues like disparity and motion are scaled to give metric depth estimates (possibly biased or unbiased, but always in a physically meaningful, metric unit of depth); extensions to accommodate correlated noise across cues (Oruç, Maloney, & Landy, 2003); and a robustness mechanism in which a depth cue that is discrepant with other depth cues can be weighted less heavily at the combination stage or even discarded.
of interest relative to the fixation point, divided by the distance from the observer to the fixation point (see Figure 1).

IC assumes that we have simultaneous depth cue measurements at several different locations on an object. If the true depths at these locations are \( z_1, \ldots, z_n \), then the disparity measurements are independent samples from the random variables \( D(z_1), \ldots, D(z_n) \), and we will denote these samples by \( d_1, \ldots, d_n \). Similarly, the retinal velocity measurements are samples from \( V(z_1), \ldots, V(z_2) \), which we will denote by \( v_1, \ldots, v_m \). The goal of IC is to use these measurements to arrive at a single depth estimate for each object location. This occurs in several steps (see Figure 2).

1. Each depth cue measurement is divided by the standard deviation of the random variable it is drawn from to produce a normalized depth cue measurement: \( \bar{d}_i = d_i / \sigma_{D} \), \( \bar{v}_i = v_i / \sigma_{V} \).
2. The normalized depth cue measurements are grouped into ordered pairs \( (\bar{d}_i, \bar{v}_i) \), where each pair consists of the normalized disparity and velocity measurements at object location \( i \). This gives a two-dimensional cloud of points.
3. The first principal component \( \bar{e}_1 \) of this cloud of points is computed.
4. The dot product is taken between each ordered pair and the first principal component: \( \rho_i = (\bar{d}_i, \bar{v}_i) \cdot \bar{e}_1 \). IC postulates that depth discrimination is based on the decision variable \( \rho_i \), and that perceived depth is some monotonically increasing function of \( \rho_i \).

We will call the computation in steps 1–4 the principal component projection (PCP) algorithm. A later stage of
IC, which we will not need to consider, determines the
monotonic relationship between \( \rho_i \) and perceived depth.

One appeal of IC is that it uses quantities that can be
measured directly from retinal images, namely absolute
disparity and retinal velocity, with no need to scale the
cues into a common, physically meaningful unit before
cue combination, as in MWF. Furthermore, Domini et al.
(2006) report experiments on perceived depth that are
consistent with IC and inconsistent with MWF. They
compare depth increment detection thresholds for stimuli
on whether the baseline stimulus is defined by single or
multiple cues, a result that is not predicted by MWF.

A reformulation

In this section, we show that IC’s cue combination
algorithm (PCP) effectively combines cues in an optimal
weighted sum. The expected values of \( D(z) = \mu z + \epsilon_i \) and
\( V(z) = \omega z + \epsilon_v \) are proportional\(^1\) and so are the expected
values of \( D(z)/\sigma_D \) and \( V(z)/\sigma_V \) from which the normalized
depth cue measurements \( d_i \) and \( v_i \) are drawn:

\[
E[V(z)/\sigma_V] = (\omega/\sigma_V)(\sigma_D/\mu)E[D(z)/\sigma_D].
\]

(1)

Thus, if the normalized depth cue measurements were
noise-free, the scatterplot of ordered pairs \((d_i, v_i)\) would lie
on the line through the origin with slope \((\omega/\sigma_V)(\sigma_D/\mu)\),
and its first principal component would be the vector
\((1,(\omega/\sigma_V)(\sigma_D/\mu))\) normalized to unit length:

\[
e^{\prime}_1 = \frac{(\mu/\sigma_D, \omega/\sigma_V)}{\left((\mu/\sigma_D)^2 + (\omega/\sigma_V)^2\right)^{1/2}}.
\]

(2)

The standard deviation of both normalized depth cues is
one, so the effect of the measurement noise is simply to
place the measurements \((d_i, v_i)\) isotropically off this line,
and in the limit of a large number of samples, the first
principal component of this cloud of points is still given by
Equation 2.

According to IC, perceived depth at object location \( i \)
is a monotonic function of \( \rho_i \):

\[
\rho_i = (d_i, v_i)e^{\prime}_1
\]

(3)

\[
= (d_i/\sigma_D, v_i/\sigma_V)\frac{(\mu/\sigma_D, \omega/\sigma_V)}{\left((\mu/\sigma_D)^2 + (\omega/\sigma_V)^2\right)^{1/2}}
\]

(4)

\[
= \frac{(\mu/\sigma_D^2)d_i + (\omega/\sigma_V^2)v_i}{\left((\mu/\sigma_D)^2 + (\omega/\sigma_V)^2\right)^{1/2}}.
\]

(5)

In Appendix A, we show how to combine two cues with
different means and standard deviations in order to
maximize the SNR of the resulting decision variable.
Comparing Equations 5 and A5 shows that \( \rho_i \) is simply the
optimal weighted sum of the disparity and the velocity
cues, scaled to unit variance.

Domini et al. (2006) explain that the goal of their cue
combination algorithm is “to obtain the best possible
estimate of the affine structure of the distal depth”
(p. 1709), and they show that IC sometimes makes the
same psychophysical predictions as MWF (p. 1714), so
we do not claim to have derived an entirely unexpected
property of IC. To support their claim to have found an
optimal combination method, though, Domini et al. only
report a numerical simulation showing that PCP works
best when the measurements \( d_i \) and \( v_i \) are normalized by
\( \sigma_D \) and \( \sigma_V \). From this simulation, it is unclear whether
PCP is optimal in any broader sense, so we believe our
derivation helps to clarify exactly what PCP accomplishes.
Furthermore, we hope that highlighting IC’s
similarity to MWF will make it easier to relate IC to the
existing cue combination literature.

Although both IC and MWF compute an optimal
weighted sum, they can make different predictions about
human performance because they assume different types
of depth cues. Consider a structure-from-motion stimulus
that has zero disparity because it is shown on a computer
monitor. IC’s PCP algorithm recognizes that the disparity
cue does not covary with the motion cue and hence seems
to convey no depth information, so IC effectively assigns
zero weight to the disparity cue.\(^2\) MWF assumes that all
cues are valid depth estimates and so combines disparity
and motion cues with weights determined by the cues’
variances (unless a robustness mechanism rejects the
disparity cue as being discrepant with other depth cues).
Thus, both models compute optimal weighted sums, but
they may weight cues differently because different notions
of optimality follow from different assumptions about
individual depth cues. Similarly, as mentioned above,
Domini et al. (2006) also describe tasks where MWF and
IC make different predictions.

We reiterate that Equation 2 for the first principal
component is only asymptotically valid, in the limit of
having depth cue measurements from a large number of
object locations. Given few measurements, PCP will not
give an exactly optimal weighted sum of individual cues.
Thus, experiments that examine what weights are assigned
to disparity and motion in impoverished stimuli like very
sparse random-dot patterns might be able to test whether
the visual system uses an algorithm like PCP to calculate
optimal weighted sums.

A JND-counting theory of perceived depth

MWF assumes that disparity and motion cues give
properly scaled estimates of true depth, and this property is
preserved in the combined depth estimate by the requirement that the cue weights sum to one. IC assumes only that depth cues are proportional to true depth, generally with different constants of proportionality. Accordingly, IC is faced with the additional problem of scaling the cues to recover true depth. The first step in IC’s solution to this problem is to posit that perceived depth is a monotonic function of $\rho_i$, which as we have shown is the optimal weighted sum of individual depth cues, scaled to unit variance. It follows immediately 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).

To see why, suppose we have a disparity-defined stimulus $d_A$ and a motion-defined stimulus $v_A$, both with a perceived depth of 10 cm, and also a disparity-defined stimulus $d_B$ and a motion-defined stimulus $v_B$ with a perceived depth of 11 cm. $d_A$ and $v_A$ have the same perceived depth, so according to IC they have the same value of $\rho_i$, which we can call $\rho_A$. Similarly, $d_B$ and $v_B$ both have $\rho_i = \rho_B$. Thus, the difference in the value of $\rho_i$ between $d_A$ and $d_B$ is $\rho_B - \rho_A$, and the difference in $\rho_i$ between $v_A$ and $v_B$ is also $\rho_B - \rho_A$. The variance of $\rho_i$ is always one, so if depth JNDs are determined by the signal and the noise properties of $\rho_i$ (as assumed by Domini et al., 2006), then the number of JNDs that separate $d_A$ and $d_B$ is the same as the number that separate $v_A$ and $v_B$. For instance, if we define one JND as a separation of $k$ standard deviations in the decision variable $\rho_i$, then $d_A$ and $d_B$ are separated by $(\rho_B - \rho_A) / k$ JNDs, and so are $v_A$ and $v_B$. Thus, IC predicts that any two depth-matched pairs of stimuli are separated by the same number of depth JNDs. (Note that even without our demonstration that PCP calculates an optimal weighted sum, Domini et al.’s Equation 7, which shows that $\rho_i$ is proportional to true depth and has unit variance, implies this same conclusion.)

In classical Fechnerian theories, JNDs correspond to equal increments in subjective stimulus magnitude. In IC, JNDs correspond to equal increments in $\rho_i$, but perceived depth is an unknown monotonic function of $\rho_i$, so JNDs need not correspond to equal increments in perceived depth. Thus, IC is more akin to revisions of Fechner’s theory that retain the JND as a unit of measurement but that allow the subjective perceptual increments corresponding to JNDs to vary as a function of baseline perceptual magnitude, e.g., an auditory JND may increase loudness more for loud sounds than for faint sounds (for a discussion of these and related issues, see Krueger, 1989).

In the following experiment, we examine the relationship between perceived depth and depth JNDs in order to test IC’s prediction that they are tightly linked. We construct 3D stimuli with disparity as a depth cue and other 3D stimuli with motion as a depth cue. We match the perceived depth of each disparity stimulus to a motion stimulus. We measure the number of depth JNDs separating pairs of motion stimuli and also the number of depth JNDs separating the corresponding depth-matched pairs of disparity stimuli. If IC is correct, then the number of JNDs separating pairs of motion-defined stimuli and pairs of depth-defined disparity-defined stimuli should be the same.

**Method**

**Observers**

Four observers participated in the experiment. Three were the authors, and the fourth was a York University undergraduate who was unaware of the purpose of the experiment. All observers reported having normal or corrected-to-normal vision.

**Stimuli**

The stimuli were horizontally oriented random-dot half-cylinders (Figures 3 and 4), modelled closely on those of Domini et al. (2006). The stimuli measured 5.0 cm horizontally and vertically and subtended 2.9 degrees of visual angle at a viewing distance of 100 cm. Each half-cylinder was defined by 200 randomly placed dots, 3.6 arcmin in diameter. One axis of the elliptical cross-section of the cylinder was vertical in the frontoparallel plane and measured 5.0 cm. The other axis was along the line of sight, and its length varied from trial to trial. That is, the stimuli were horizontal half-cylinders, stretched or compressed along the observer’s line of sight to varying degrees. A vertical depth gradient was created using either binocular disparity or motion. In the motion condition, a percept of depth was created by rocking the cylinder sinusoidally about a horizontal axis of rotation through the nearest point of the cylinder. The oscillation frequency was 1.0 Hz, and the amplitude was 10 degrees. All stimuli were shown for 1000 ms.

Stimuli were shown on a Dell UltraScan P991 19-in. CRT display with a 1024 × 768 resolution and a 120-Hz refresh rate. The display measured 34.5 × 26.0 cm and each pixel measured 0.35 mm. Stimuli were generated and displayed using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) on an Apple G5 computer. All stimuli were viewed binocularly through Stereographics CrystalEyes LCD-shuttered glasses at a refresh rate of 60 Hz in each eye.

**Procedure**

**Depth matching**

In the depth-matching part of the experiment, we matched the perceived depth of half-cylinders defined by disparity to half-cylinders defined by motion, individually for each observer. As reference stimuli, we used...
motion-defined half-cylinders with simulated depths of 1.25 cm, 2.5 cm, and 5.0 cm. The vertical extent of the half-cylinders was 5.0 cm, so these stimuli were flatter than circular, circular, and deeper than circular, respectively. As test stimuli, we used disparity-defined half-cylinders with a range of simulated depths. Observers completed three blocks of 270 two-interval forced-choice (2IFC) trials, with only a single motion-defined reference depth shown in each block. On each trial, observers viewed a motion-defined reference stimulus and a disparity-defined test stimulus, in random order, separated by a blank interstimulus interval of 750 ms and pressed a key to indicate which interval contained the stimulus with the greater perceived depth. No feedback was given. The simulated depth of the test stimulus was chosen by the method of constant stimuli from a set of nine depths ranging from zero to twice the depth of the reference stimulus assigned to the block. All stimuli were viewed through LCD-shuttered stereo glasses, but the reference stimuli had zero disparity.

For each reference stimulus, we calculated the psychometric function indicating the probability of the observer choosing the disparity-defined test stimulus as having the greater perceived depth as a function of the simulated depth of the test stimulus. We fitted a Weibull cumulative distribution function to each psychometric function, and we took the fitted 50% probability point as the point of subjective equality, i.e., the simulated depth at which the disparity-defined test stimulus had the same perceived depth as the motion-defined reference stimulus.

**Depth discrimination**

In the discrimination part of the experiment, we measured JNDs for the three motion-defined reference stimuli and the three depth-matched disparity-defined stimuli determined in the first part of the experiment. We call these the six *matched* stimuli. We measured the six JNDs in separate 270-trial blocks, i.e., only a single matched stimulus was shown in a given block. On each 2IFC trial, observers viewed a matched stimulus and a test...
stimulus, with the simulated depth of the test stimulus chosen by the method of constant stimuli. Disparity-defined matched stimuli were shown with disparity-defined test stimuli, and motion-defined matched stimuli were shown with motion-defined test stimuli. The stimuli were shown for 1000 ms, in random order, separated by a blank 750-ms interstimulus interval, and the observer pressed a key to indicate which interval contained the stimulus with the greater perceived depth.

We defined a JND as the difference between the depths at which observers achieved 50% and 75% correct performance.

Results and discussion

Depth matching

Figure 5 shows a depth matching psychometric function for a typical observer. As expected, the probability of the observer reporting that the disparity-defined test stimulus had a greater perceived depth than the motion-defined reference stimulus increased smoothly as a function of the simulated depth of the test stimulus.

Figure 6 shows depth matches for all observers and indicates that there was an approximately linear relationship between the simulated depth of disparity-defined and motion-defined stimuli at the point of subjective equality. Furthermore, the disparity-defined stimuli had less simulated depth than the motion-defined stimuli at points of subjective equality, meaning that at a given simulated depth, motion-defined stimuli were perceived as being shallower than disparity-defined stimuli. This may be because the motion-defined stimuli had zero disparity, which was a cue that the stimuli were actually flat. In any case, this is not a problem for our test of IC: All we need are two physically different sets of stimuli with matched perceived depths, and this is what the depth-matching part of the experiment provided.

Depth discrimination

Figure 7 shows depth JNDs for all observers. JNDs increased approximately linearly as a function of stimulus depth, meaning that depth discrimination performance roughly followed Weber’s law over the stimulus range tested (as found by McKee, Levi, & Bowne, 1990, for disparity-defined stimuli). This finding is itself important for our test of IC. According to IC, discrimination performance is based on the random variable \( \rho \), which is proportional to true depth and has a variance that is independent of \( z \). Thus, IC assumes that JNDs are the same at all depths, but this is clearly not the case.

If JNDs were the same at all depths, we could count the number of JNDs separating two disparity-defined stimuli simply by dividing their simulated depth difference by the unique JND for disparity, and we could make the corresponding calculation for motion-defined stimuli. Now, though, we find that there is a different JND for each stimulus, so the simple calculation suggested by IC is not an appropriate way of counting JNDs.

Nevertheless, to remain true to the original formulation of IC, we performed this analysis. Figure 8 shows the number of JNDs separating disparity-defined and motion-defined pairs of stimuli, estimated using the JND from the shallower stimulus in each pair, and Figure 9 shows the same calculation but using the JND from the deeper stimulus in each pair. These figures indicate that depth-matched stimulus, with the simulated depth of the test stimulus shown with disparity-defined test stimuli, and motion-defined matched stimuli were shown with motion-defined test stimuli. The stimuli were shown for 1000 ms, in random order, separated by a blank 750-ms interstimulus interval, and the observer pressed a key to indicate which interval contained the stimulus with the greater perceived depth. No feedback was given.

We defined a JND as the difference between the depths at which observers achieved 50% and 75% correct performance.

Figure 5. A psychometric function for a typical observer in the depth-matching condition. The reference stimulus was a motion-defined half-cylinder with a simulated depth of 2.5 cm (vertical dotted line). The x-axis indicates the simulated depth of the disparity-defined test stimulus, and the y-axis shows the proportion of times the observer chose the test stimulus as having the greater perceived depth.

Figure 6. Depth matches for all observers. The x-axis indicates the simulated depth of the motion-defined reference stimulus, and the y-axis shows the simulated depth of the disparity-defined test stimuli at the point of subjective equality. The dotted diagonal line shows where depth matches would fall if motion-defined and disparity-defined stimuli with equal simulated depths also had equal perceived depths.
pairs of stimuli were not always separated by equal numbers of JNDs, contradicting IC’s prediction. However, these results should be regarded with caution because depth discrimination performance approximately followed Weber’s law, so the calculation suggested by IC is not a valid way of counting the number of JNDs separating two stimuli.

The claim that perceived depth is measured out in JNDs is an interesting one, however, so to test this possibility we recalculated the number of JNDs separating the pairs of depth-matched stimuli in our experiment, this time taking into account Weber’s law. If the depth JND for a disparity-defined stimulus is proportional to the baseline stimulus depth, \( JND_D(z) = k_D z \), then the number of JNDs separating disparity-defined stimuli at depths \( z_1 \) and \( z_2 \) is

\[
 n_D = \int_{z_1}^{z_2} \left( \frac{1}{JND_D(z)} \right) dz = \int_{z_1}^{z_2} \left( \frac{1}{k_D z} \right) dz
\]

\[
 = \left( \frac{1}{k_D} \right) \left( \ln |z_2| - \ln |z_1| \right),
\]

where \( \ln \) is the natural logarithm. Similarly, the number of JNDs separating motion-defined stimuli at depths \( z_1 \) and \( z_2 \) is

\[
 n_M = \left( \frac{1}{k_M} \right) \left( \ln |z_2| - \ln |z_1| \right),
\]

where \( k_M \) is the constant of proportionality in Weber’s law for motion-defined stimuli, \( JND_M(z) = k_M z \).

Figure 10 shows the number of JNDs separating pairs of motion-defined and disparity-defined stimuli, calculated using the JND from the shallower of each stimulus pair. In each panel, the leftmost bar shows the number of JNDs separating the 1.25-cm and the 2.5-cm motion-defined stimuli, and the immediately adjacent bar shows the number of JNDs separating the disparity-defined stimuli that were depth-matched to those two motion stimuli. The next pair of bars shows the number of JNDs between the 1.25-cm and the 5.0-cm motion stimuli and the corresponding depth-matched disparity stimuli. The third pair of bars shows the number of JNDs between the 2.5-cm and the 5.0-cm motion stimuli and the corresponding depth-matched disparity stimuli. The error bars indicate 95% confidence intervals, and asterisks indicate significantly different motion and disparity JND counts (\( p < 0.05 \), based on two-tailed bootstrap tests). Note that these JND counts are calculated using the formula derived from IC, which mistakenly assumes that JND size is independent of baseline stimulus depth. Hence, these JND counts are suspect.
using Equations 6 and 7. We calculated the constants $k_D$ and $k_M$ individually for each observer by making a maximum-likelihood linear fit to JND size versus simulated depth (as in Figure 7, but using each observer’s JNDs instead of the group means). Even this revised calculation, which takes into account Weber’s law and thus gives a more accurate JND count, indicates that depth-matched motion and disparity stimuli were not separated by the same number of JNDs. In every comparison, the JND count was less for motion-defined stimuli than for the corresponding disparity-defined stimuli. Not all of the JND counts shown in Figure 7 are independent, as the three motion JND counts are calculated from all three possible pairings of the three motion stimuli and similarly for the disparity JND counts. Nevertheless, even if we just consider the 1.25-cm vs. the 2.5-cm pairs and the 2.5-cm vs. the 5.0-cm pairs, this means that in eight cases the motion JND count was less than the disparity JND count, which is a statistically significant difference under a sign test ($p < 0.01$).
Furthermore, several of the individual JND count comparisons were statistically significant as well ($p < 0.05$). Thus, a key psychophysical prediction of IC is incorrect, even when we use a JND-counting formula that takes account of Weber’s law.

**Conclusion**

As far as we know, this experiment is the first test of a JND-counting model of perceived depth. Tests 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 JNDs does not predict differences in sensory magnitudes (e.g., Newman, 1933; Stevens, 1961). Our results are consistent with this literature. However, given the number of proposed extensions and revisions of Fechner’s account (see Krueger, 1989), a complete review of this issue is beyond the scope of this paper.

Domini et al. (2006) argue that their own empirical findings support IC and cannot be reconciled with MWF. We do not take issue with their conclusions, which address different aspects of IC than those we have discussed, e.g., they argue that IC correctly predicts that observers are largely unable to make metric depth estimates and only recover useful depth information up to an arbitrary affine transform. Many others have investigated similar claims (see Todd, 2004), and it is not our intent to review them here. Rather, our aim has been to investigate other implications of the IC model. In sum, we find that the separation between two objects as measured in JNDs does not predict their separation in perceived depth. This finding can be accommodated by theories like MWF that allow perceived depth and depth discriminability to vary independently, but it is problematic for theories like IC that imply that perceived depth and depth discriminability are closely linked.

**Appendix A**

Given two Gaussian random variables, $S \sim N(\mu_S, \sigma_S^2)$ and $T \sim N(\mu_T, \sigma_T^2)$, what weights maximize the SNR of the weighted sum $C = \mu S + \nu T$?

The mean of the weighted sum is $u \mu_S + v \mu_T$, and the variance is $u^2 \sigma_S^2 + v^2 \sigma_T^2$, so the SNR is $(u \mu_S + v \mu_T)/\left(u^2 \sigma_S^2 + v^2 \sigma_T^2\right)^{1/2}$. Any two pairs of weights $(u, v)$ and $(k_u, k_v)$ that differ only by a scale factor give the same SNR, so we will add the constraint that the variance of the weighted sum is one, $u^2 \sigma_S^2 + v^2 \sigma_T^2 = 1$. (The MWF model assumes $u + v = 1$, but the unit-variance constraint is more useful in our discussion of IC.)

We will use the method of Lagrange multipliers (Byron & Fuller, 1992). The objective function is the SNR, $f(u, v) = (u \mu_S + v \mu_T)/\left(u^2 \sigma_S^2 + v^2 \sigma_T^2\right)^{1/2}$, subject to the unit-variance constraint $g(u, v) = u^2 \sigma_S^2 + v^2 \sigma_T^2 - 1 = 0$. The Lagrangian is

$$L(u, v, \lambda) = f(u, v) + \lambda g(u, v),$$

$$= (u \mu_S + v \mu_T)/\left(u^2 \sigma_S^2 + v^2 \sigma_T^2\right)^{1/2} + \lambda(u^2 \sigma_S^2 + v^2 \sigma_T^2 - 1).$$

Setting $\nabla L = 0$ and solving for $u$ and $v$, we find

$$u = (\mu_S/\sigma_S^2)/\left((\mu_S/\sigma_S^2)^2 + (\mu_T/\sigma_T^2)^2\right)^{1/2},$$

$$v = (\mu_T/\sigma_T^2)/\left((\mu_S/\sigma_S^2)^2 + (\mu_T/\sigma_T^2)^2\right)^{1/2}.$$  

Thus, the optimal unit-variance weighted sum is

$$C = \frac{(\mu_S/\sigma_S^2)S + (\mu_T/\sigma_T^2)T}{\left((\mu_S/\sigma_S^2)^2 + (\mu_T/\sigma_T^2)^2\right)^{1/2}}.$$  

**Footnotes**

1. This proportionality could be broken by creating cue conflict stimuli where disparity and motion specify different affine structures, and the analysis that follows does not apply in such unusual cases. Most cue conflict stimuli used to date, however, have specified the same affine depth structure in all cues, and have just assigned different depth scale factors to different cues.

In this case, the principal component in Figure 2 will be vertical, and so the dot product $\rho_i = \langle \vec{d}_i, \vec{e'}_i \rangle$ will be independent of $\vec{d}_i$. In order to accommodate Weber’s law, IC would have to be revised to change the signal and noise properties of

Acknowledgments

We thank Fulvio Domini and Michael Landy for helpful discussions. This research was funded by grants from the Natural Sciences and Engineering Research Council to RFM and LMW.

Commercial relationships: none.

Corresponding author: Kevin J. MacKenzie.

Email: k.j.mackenzie@bangor.ac.uk.

Address: School of Psychology, Adeilad Brigantia, Bangor University, Bangor, Gwynedd, LL57 2AS, UK.
the decision variable. Domini and Caudek (2007) have
started investigations along these lines. However, incor-
porating Weber’s law simply by making each cue’s
standard deviation proportional to the value of the cue.
will not work because then all normalized disparity
measurements \( \tilde{d}_i \) have the same expected value, as do all
normalized velocity measurements \( \tilde{v}_i \), and the decision
variable \( \rho \) is independent of true depth.

References


classical and quantum physics. New York: Dover
Publications, Inc.

approach to the problem of cue integration. European
Conference on Visual Perception, Arezzo, Italy.

and motion information are not independently pro-
cessed by the visual system. Vision Research, 46,
1707–1723. [PubMed]

Ernst, M. O., & Banks, M. S. (2002). Humans integrate
visual and haptic information in a statistically optimal

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S.
(2004). Slant from texture and disparity cues: Optimal
[PubMed] [Article]

Jacobs, R. A. (1999). Optimal integration of texture and
motion cues to depth. Vision Research, 39, 3621–3629.
[PubMed]

Integration of stereopsis and motion shape cues.
Vision Research, 34, 2259–2275. [PubMed]

Krueger, L. E. (1989). Reconciling Fechner and Stevens:
Toward a unified psychophysical law. Behavioral and
Brain Sciences, 12, 251–320.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M.
combination: In defense of weak fusion. Vision

framework for robust fusion of depth information. In
W. A. Pearlman (Ed.), Visual communications and
image processing: IV. Proceedings of the SPIE (vol.
1199, pp. 1154–1163).

imprecision of stereopsis. Vision Research, 30,
1763–1779. [PubMed]

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.

Weighted linear cue combination with possibly
[PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual
psychophysics: Transforming numbers into movies.

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]

Trends in Cognitive Sciences, 8, 115–121. [PubMed]

A perturbation analysis of depth perception from
combinations of texture and motion cues. Vision
Research, 33, 2685–2696. [PubMed]