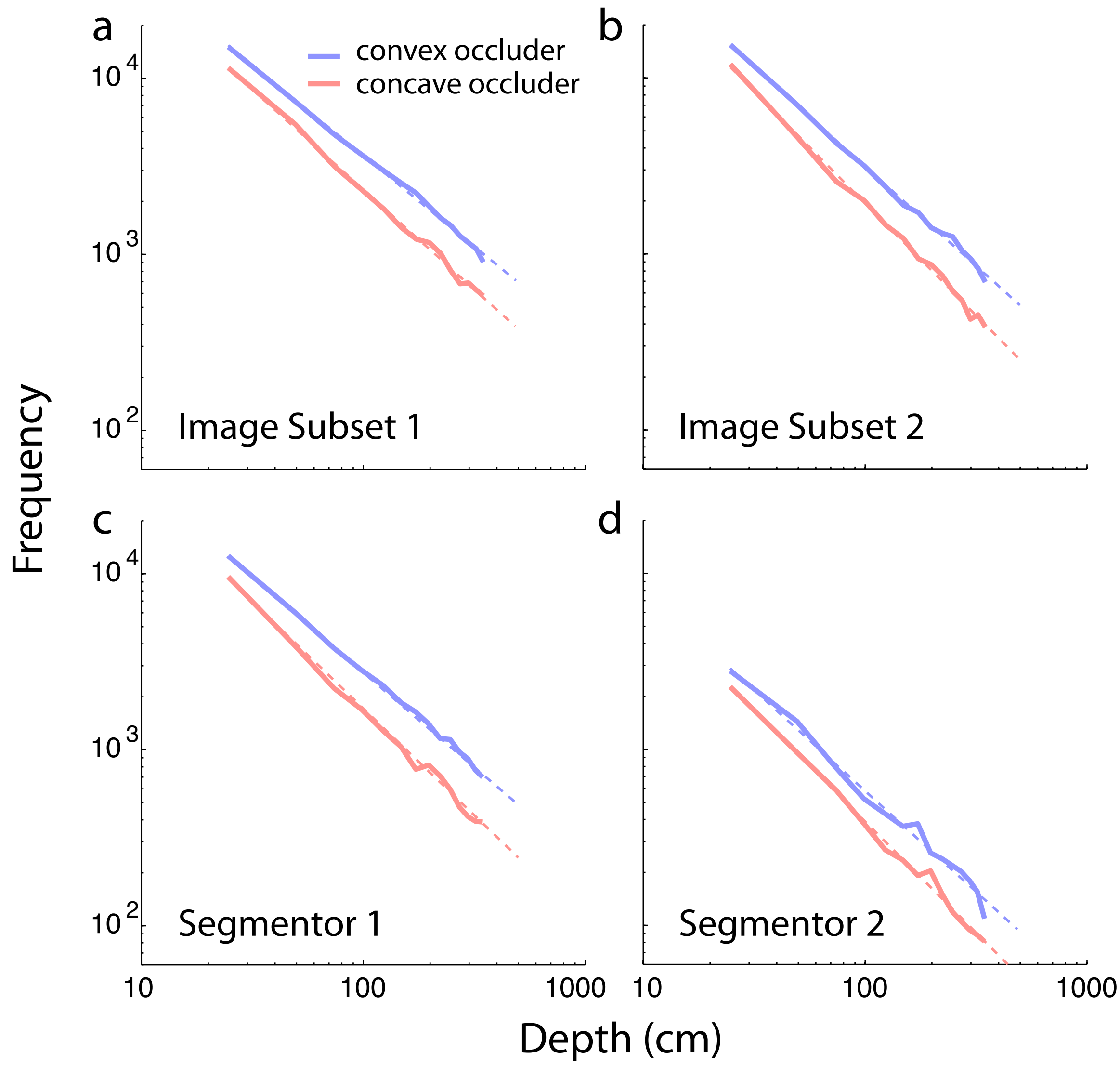

Figure S1: Robustness of power law relationship between convexity and depth. Frequency of occurrence is plotted as a function of depth for all convexities greater than zero (light blue) and all convexities less than zero (pink) for different subsets of images and segmenters. Dotted lines are power-law fits to the data. **a)** Convexity-depth relationship for a random sampling of half of the segmentations. **b)** Convexity-depth relationship for the other half of the segmentations. **c)** Convexity-depth relationship for the human segmenter who segmented the most images (27). **d)** Convexity-depth relationship for another segmenter who segmented only nine images.

Cue integration with intrinsic and extrinsic uncertainty

Johannes Burge, Charless C. Fowlkes, Martin S. Banks

An ideal observer model with extrinsic and intrinsic uncertainty

To understand how information from various visual cues should be integrated, we need to distinguish two types of uncertainty. *Intrinsic uncertainty* is due to properties of the visual system such as noise in cue measurement and in cue interpretation. *Extrinsic uncertainty* is due to uncertainty in the relation between the proximal stimulus and the scene property of interest. This type of uncertainty is captured by natural-scene statistics. If an ideal observer is repeatedly presented with the same stimulus, intrinsic noise will make the observer's response stochastic. Extrinsic uncertainty, on the other hand, should not contribute to response stochasticity because the stimulus is held fixed.

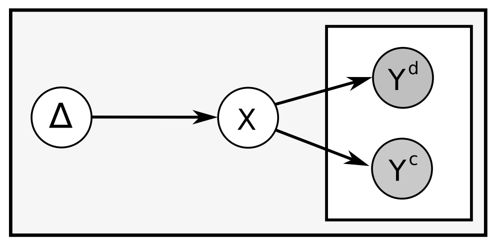

Figure S2: A graphical model of cue combination with extrinsic $P(X|\Delta)$ and intrinsic $P(Y^d, Y^c|X)$ uncertainty.

Consider the depth cues in our experiment: disparity and convexity. In normal viewing conditions, disparity is subject to intrinsic uncertainty (e.g., noise in disparity measurement and noise in signals, like eye position, necessary for mapping disparity into depth), but very little extrinsic uncertainty because the mapping from depth to presented disparity is deterministic (i.e. presented disparity depends only on depth and viewing geometry). Convexity, on the other hand, is subject to both intrinsic uncertainty (e.g., noise in convexity measurement) and extrinsic uncertainty because the mapping from depth to presented convexity is non-deterministic (image region convexity depends not just on depth and viewing geometry, but on the projected shape of the occluder). In the case of convexity, extrinsic uncertainty vastly exceeds intrinsic uncertainty such that any noise in the convexity measurement will have negligible effect on the estimated depth (because small changes in convexity have essentially no effect on the convexity-depth probability distribution).

We can express this more precisely in terms of random variables shown in Figure S2. Let Δ be the scene depth to be estimated by the observer, X the retinal stimulus (image) and Y^d and Y^c the cue measurements of disparity and convexity available to the ideal observer. Sample a natural scene from the world. A particular depth Δ is related to the image of the scene X by the distribution $P(X|\Delta)$ which incorporates the statistical relation between images and scene geometry. Given the retinal stimulus X, the observer computes cues Y^d and Y^c and uses them to estimate the value of Δ. When there is intrinsic noise, these computations are subject to uncertainty. For example, previous research suggests that the disparity cue is subject to zero-mean, additive Gaussian noise η whose variance is a function of X, that is, $Y^d = g(X) + \eta$ where g is some function that extracts the disparity cue from the sensory input and $\eta \sim N(0, \sigma^2_{dint}(X))$. The outer box in Figure S2 corresponds to sampling scenes while the inner box indicates repeated trials where the scene Δ and stimulus X are held fixed and the cues Y are recomputed.

Optimal decisions

In a 2IFC experiment, an ideal observer compares two different stimuli and judges if the first or second has greater depth. In our analysis, we assume the posterior distribution over Δ given the measurements Y^d, Y^c is well approximated by a Gaussian with parameters μ_1, σ_1^2 and μ_2, σ_2^2 where these parameters are a function of the particular cue measurements y_1^d, y_1^c and y_2^d, y_2^c. Assuming that different error types incur equal costs, the optimal response is:

$$\begin{aligned}
R(y_1^d, y_1^c, y_2^d, y_2^c) &= P(\Delta_1 > \Delta_2 | y_1^d, y_1^c, y_2^d, y_2^c) > 0.5 \\
&= P(\Delta_1 - \Delta_2 > 0 | y_1^d, y_1^c, y_2^d, y_2^c) > 0.5 \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(x-(\mu_1-\mu_2))^2}{\sigma_1^2+\sigma_2^2}} dx > 0.5 \\
&= 1 - \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) > 0.5 \\
&= (\mu_1 - \mu_2) > 0
\end{aligned}$$

where Φ is the Gaussian cumulative distribution function and we have used the fact that the difference of two Gaussian variables is Gaussian. This gives a simple decision rule: compare the estimates of the posterior means and report which was larger. This rule also holds more generally so long as the posterior distribution is symmetric and unimodal. The ideal observer's response is therefore a deterministic function of the measured cues.

Repeated trials

If there is no intrinsic uncertainty, repeated trials with the same stimulus X yield the same Y and hence the same response R. When intrinsic uncertainty is present, the values of Y_1 and Y_2, and hence μ_1, μ_2, fluctuate. This results in variable responses for repeated presentations of the identical pair of stimuli x_1 and x_2. To understand the effect of these fluctuations, we compute the expected value of R across repeated experiments where the stimuli x_1, x_2 are fixed.

We derive an expression for $E[R]$ assuming that the cues are conditionally independent and yield Gaussian likelihood distributions over Δ centered at values μ_d and μ_c with variances σ_d^2 and σ_c^2. In general, only part of the total variance of the likelihoods σ_d^2, σ_c^2 is due to intrinsic noise; the remainder is due to extrinsic sources. Let us assume the intrinsic noise component is zero-mean Gaussian with variances $\sigma^2_{dint}, \sigma^2_{cint}$.

For conditionally independent Gaussian cues, the ideal observer combines the cues to get a posterior distribution over Δ with mean

$$\mu(Y^d, Y^c) = \frac{\sigma_c^2 \mu_d + \sigma_d^2 \mu_c}{\sigma_d^2 + \sigma_c^2}$$

In repeated trials, the estimate $\mu(y^d, y^c)$ is a linear function of two independent Gaussian random variables and is hence itself Gaussian with the same mean $m = \mu(y^d, y^c)$ and variance

$$s^2 = \left(\frac{\sigma_c^2}{\sigma_d^2 + \sigma_c^2}\right)^2 \sigma^2_{dint} + \left(\frac{\sigma_d^2}{\sigma_d^2 + \sigma_c^2}\right)^2 \sigma^2_{cint}$$

With the distribution of μ in hand we can compute the expected response as

$$\begin{aligned} E[R] &= \int (\mu_1 - \mu_2 > 0) p(y_1^d, y_1^c | x_1) p(y_2^d, y_2^c | x_2) dy_1^d dy_1^c dy_2^d dy_2^c \\ &= P(\mu_1 - \mu_2 > 0 | x_1, x_2) \\ &= \Phi \left(\frac{m_2 - m_1}{\sqrt{s_1^2 + s_2^2}} \right) \end{aligned}$$

To generate a psychometric curve for the ideal observer, we hold x_1 fixed as a standard, adjust the comparison x_2, and compute the expected response for each pairing x_1, x_2. In general, this psychometric curve will not be the CDF of a Gaussian because the variance s_2^2 in the denominator is a function of x_2.

Two specific cases of interest are:

- **Purely intrinsic uncertainty:** If we assume that the uncertainty in the likelihoods is due entirely to intrinsic sources, then $\sigma_{dint}^2 = \sigma_d^2$ and $\sigma_{cint}^2 = \sigma_c^2$. In this case, the expression for the variance in the estimate simplifies to

$$s^2 = \frac{\sigma_d^2 \sigma_c^2}{\sigma_d^2 + \sigma_c^2}$$

  so that the response function and the expected response have exactly the same form.

- **Mixed intrinsic and extrinsic uncertainty:** If we assume that Y^d suffers only intrinsic noise while Y^c is subject to only extrinsic uncertainty, then $\sigma_{dint}^2 = \sigma_d^2$ and $\sigma_{cint}^2 = 0$, so the mean estimate m will remain the same as before, but the second term in the expression for the variance vanishes leaving

$$s^2 = \left(\frac{\sigma_c^2}{\sigma_d^2 + \sigma_c^2} \right)^2 \sigma_d^2$$

  Because the (extrinsic) variance of the convexity cue is quite large relative to the (intrinsic) variance of the disparity cue, the fraction is close to 1 so $s^2 \approx \sigma_d^2$ and thus the slope of the psychometric function is determined by σ_d^2.

In fitting our model (see Methods), the convexity-depth distribution is not assumed to be Gaussian but the posterior distribution is still well approximated by a Gaussian with mean $\mu(Y^d, Y^c) = \mu_d + m_d \sigma_d$ and variance σ_d^2 where m_d is the slope of the convexity-disparity distribution associated with the convexity cue. We make the reasonable assumption that the intrinsic uncertainty in convexity is small enough that there is little variation in the slope m_d during repeated trials. Thus, the variance of the posterior mean $\mu(Y^d, Y^c)$ and hence the slope of the psychometric curve produced by the model is determined by $\sigma_d^2 = \sigma_{dint}^2$. As a consequence, our model fitting methodology is perfectly appropriate for these particular conditions.