In press, *Psychological Review*

Information along contours and object boundaries

Jacob Feldman and Manish Singh
Department of Psychology, Center for Cognitive Science
Rutgers University—New Brunswick

Attneave (1954) famously suggested that information along visual contours is concentrated in regions of high magnitude of curvature, rather than being distributed uniformly along the contour. Here we give a formal derivation of this claim, yielding an exact expression for information, in Shannon's sense, as a function of contour curvature. Moreover, we extend Attneave's claim to incorporate the role of *sign* of curvature, not just *magnitude* of curvature. In particular, we show that for closed contours, such as object boundaries, segments of negative curvature (that is, concave segments) literally carry greater information than corresponding regions of positive curvature (i.e., convex segments). The psychological validity of our informational analysis is supported by a host of empirical findings demonstrating the asymmetric way in which the visual system treats regions of positive and negative curvature.

In 1954, Attneave proposed that information along a visual contour is concentrated in regions of high magnitude of curvature, rather than distributed uniformly.[1] His observation was informal, but astute, and helped to inspire interest in information-processing approaches to the study of vision. Fig. 1a shows a shape with points of locally maximal magnitude of curvature marked. By way of demonstration that such points convey most of the psychologically important information about shape, Attneave drew a line drawing of a cat by taking only the points of local maxima of curvature magnitude, and joining them with straight line segments.[2] The resulting line drawing (now popularly known as 'Attneave's cat') was easily recognizable, suggesting that not much loss of information had occurred. Attneave (1954) also briefly described the results of an experiment in which participants were asked to approximate two-dimensional shapes with a fixed number of points, and then asked to indicate where on the original shapes these points were located. Histogram plots of the points selected revealed salient peaks at precisely the points of local maxima of curvature magnitude (similar

Supported in part by NSF SBR-9875175 to J.F. and NSF BCS-0216944 to M.S. We are very grateful to Michael Kubovy, Howard Resnikoff, Chester Stromswold, and two anonymous reviewers for thoughtful comments on the manuscript. Please direct correspondence to Jacob Feldman, Dept. of Psychology, Center for Cognitive Science, Rutgers University – New Brunswick, 152 Frelinghuysen Rd., Piscataway, New Jersey, 08854, or by e-mail at jacob@ruccs.rutgers.edu.

[1] Curvature is sometimes treated as an *unsigned* quantity—the magnitude of the tangent deriviative or the "degree of bendiness"—and sometimes as a *signed* quantity, in which case sign is conventionally assigned positive for turns towards the interior of the "figure" (i.e., on convexities) and negative for turns away from the interior (i.e., in concavities). These discrepant senses can cause confusion, for example when "low curvature" can refer either to a relatively straight curve (when curvature is used in the unsigned sense) or a region with high magnitude in the negative direction (i.e. a sharp concavity). Attneave used the term curvature in its unsigned sense. Thus in modern language his claim was that information depends on the *magnitude* of curvature. He made no reference in his paper to the sign of curvature, and his proposal did not distinguish between convex and concave regions of a contour.

[2] Irving Biederman (speaking informally at the Psychonomic Society conference, November, 2000) pointed out that in Attneave's own telling—and contrary to myth—Attneave never actually made a smoothly curved line drawing of a cat. Rather, Attneave drew the famous feline polygon by hand directly from visual inspection of his own pet.

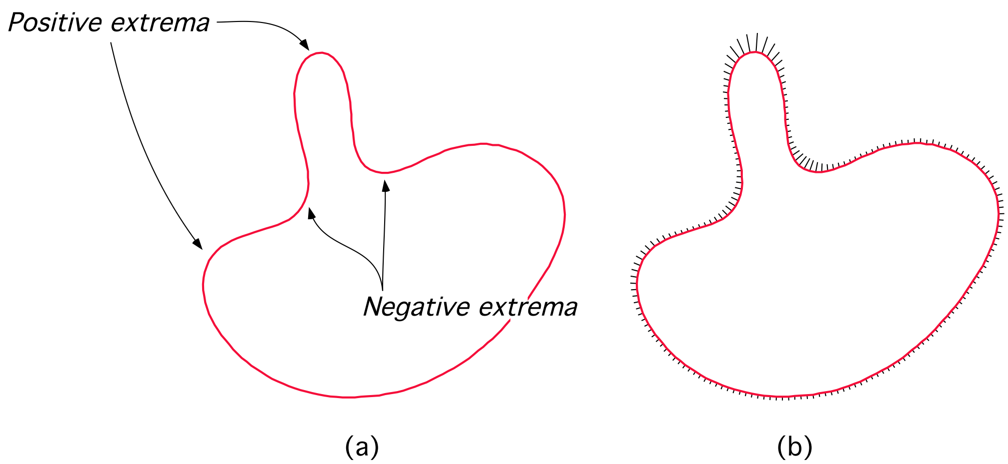

Figure 1. Information on the boundary of a shape is concentrated in regions of high magnitude of curvature. (a) A shape with curvature extrema marked, including both positive (convex) extrema and negative (concave) extrema (i.e., minima of signed curvature). (b) The same shape with contour information (surprisal) plotted (Eq. 4), reminiscent of Attneave's (1954) histograms.

to Fig. 1b). The details of Attneave's experiment were apparently never published (Attneave's 1954 article cites only a 'mimeographed note'). However Norman, Phillips, and Ross (2001) have recently conducted a similar experiment and replicated Attneave's results. Moreover, contour deletion experiments (Biederman and Blickle, discussed in Biederman, 1987) have shown that deletion of high-curvature contour segments creates greater difficulties in recognition than deletion of low-curvature segments of comparable length, demonstrating the special role high-curvature contour segments play in recognition.

Resnikoff (1985) has provided a derivation of Attneave's claim, based on Shannon's mathematical definition of information. Although Resnikoff deserves credit for placing Attneave's proposal on a formal footing for the first time, we feel that his derivation has several problems that leave it short of providing a mathematical substantiation of Attneave's idea (see Appendix). In this article, we provide a novel derivation of the information content of contours, which does not require the assumptions implicit in Resnikoff's analysis, but rather is informed by recent psychophysical findings about the mental representation of curves. Moreover, we extend the informational analysis to the case of *closed contours*—as might correspond to object boundaries—deriving an asymmetry in the information content of negative and positive curvature regions. This analysis extends Attneave's original claim—which treats positive and negative curvature regions symmetrically—and is supported by a host of empirical findings in the literature demonstrating the influence of sign of curvature on shape perception.

Information

We begin with a statement of Shannon's (1948) formula for a continuous measure M. Assume first a distribution (probability density function) $p(M)$, which represents the observer's beliefs about the value of M before a measurement is taken. What information is gained by measuring M? Shannon's insight was that this depends on the value obtained, and, more specifically, on its likelihood. If the observed value of M is relatively close to what was expected—say, it was the most likely case—then relatively little information has been gained by measuring it. But if it reveals a surprising value—say, something in the tails of the distribution $p(M)$—then relatively much information has been gained. Specifically, Shannon showed that this dependence must follow the negative logorithm[3] of the probability, i.e.,

$$u(M) = -\log[p(M)]. \tag{1}$$

The quantity $u(M)$ is sometimes called the *surprisal* of M. The *information* contained in the distribution $p(M)$, i.e. the entire ensemble of probabilities $p(M)$ taken as a whole, is simply the expected value of the surprisal,

$$I(p) = -\sum_M p(M)\log[p(M)]. \tag{2}$$

that is, the mean of the all possible surprisals weighted by their probabilities.

Contours

Now consider the case of a simple planar curve (i.e., with no self-intersections) of length L, sampled at n uniformly-spaced points separated by intervals $\Delta s = L/n$ (Fig. 2). From point to point along the sampled curve, the tangent direction changes by an angle $\Delta\phi$ or α, called the *turning angle.* (Without loss of generality, we assign the field of normals such that positive values of $\Delta\phi$ correspond to clockwise turns, and negative values to counter-clockwise turns.) Turning angles, and relative orientations between edge pairs more generally, are a parameter of widespread interest in the vision literature,

[3] Treatments of information theory usually assume logs in base 2, but the choice of base does not really matter since the resulting quantities differ only by a multiplicative constant. In what follows we actually use base e for reasons that will become apparent.

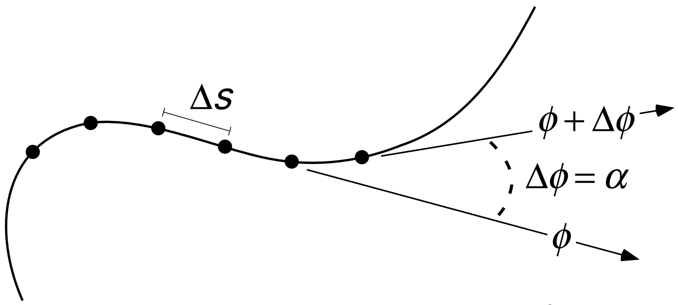

Figure 2. A simple plane curve sampled at intervals of arclength Δs. Each point has a tangent ϕ; the angle $\Delta\phi$ between successive tangents is denoted α.

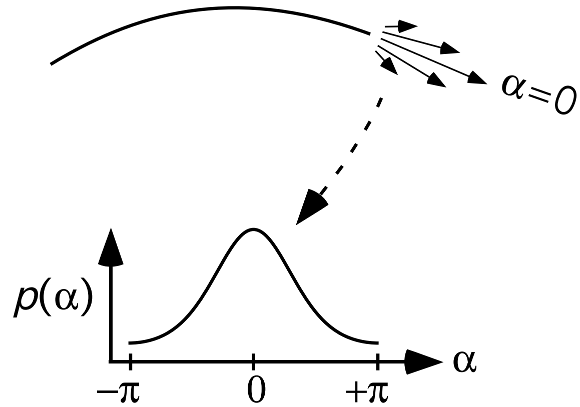

Figure 3. The expected change in tangent direction α is distributed as a von Mises distribution centered on 0 (straight).

for example in the enormous literatures on contour integration in both psychophysics (Caelli & Umansky, 1976; Field, Hayes, & Hess, 1993) and physiology (Bosking, Zhang, & Fitzpatrick, 1997; Gilbert, 1995). The reasons for this interest are very basic. Absolute position is very unlikely to be a parameter of interest to the visual system, because of the need for translation invariance; and likewise for absolute orientation because of the (perhaps weaker) need for rotational invariance (see Jaynes, 1973). Hence turning angle, which is invariant to both translation and rotation, is a more plausible choice as a parameter of a priori interest.

Now, having chosen turning angle α as a parameter of interest, in what way are specific observed values of α informative? This depends on the assumed distribution $p(\alpha)$ (see Fig. 2). That is, as one moves around the curve, choosing successively the next change in tangent direction, from what distribution are these choices drawn?

In what follows we will assume that the change in tangent direction on a smooth curve follows a von Mises distribution centered on "straight" $\alpha = 0$,

$$p(\alpha) = A \exp[b \cos(\alpha)]. \qquad (3)$$

where b is a parameter[4] modulating the spread of the distribution (acting like the inverse of variance), and A is a normalizing constant (depending on b but not on α). The von Mises distribution (due to von Mises, 1918) is the natural analog of a Gaussian (normal) distribution for the case of angular measurements[5] (see Fisher, 1993). It serves as an appropriate choice for a distribution of tangent directions for several reasons. First, like the Gaussian in the case of non-angular measurements, it maximizes entropy given a fixed mean and spread (Mardia, 1972), meaning here that it represents in a maximally neutral manner the assumption that the tangent will have a particular expected direction and a particular magnitude of uncertainty around that direction. Second, it agrees with a variety of empirical data, including the tuning curves of orientation-selective neurons in the primary visual cortex (Swindale, 1998), as well as human observers' subjective expectations about how smooth curves are likely to continue (Feldman, 1995, 1996, 1997, 2001).[6] The precise choice of distribution is actually not very important to our argument (see below for substantiation of this claim); the important properties of this choice of distribution are that (i) it is centered at $\alpha = 0$, meaning that straight continuation of the tangent direction is considered the most likely case, and (ii) probability decreases symmetrically with deviations from straight. These properties are indeed observed in the actual statistics of orientation changes at successive points along perceptual contours in natural images (Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001).

Now, at a particular point along the curve, and particular choice of angle α, what is the information at that point? Following Shannon, all we can give for a particular measurement is its surprisal. Combining Eqs. 1 and Eq. 3, we get

$$u(\alpha) = -\log[p(\alpha)] = -\log A - b\cos(\alpha) \qquad (4)$$

The first term is an additive constant, not dependent on α, which gives the absolute minimal surprisal, obtained in the case of a straight line; its exact value derives from the specifics of the von Mises distribution. The second term, $-b\cos(\alpha)$, shows how the suprisal depends on α: as the negative cosine, which means it increases monotonically with increasing deviations from straight.[7] Fig. 1b shows a plot of

---

[4] The symbol κ is often used to denote the spread parameter b, but we reserve κ for curvature, used below.

[5] A Gaussian is not well-defined for angles because its support is $(-\infty, \infty)$ while angles are only defined in $(-\pi, \pi)$. The von Mises asymptotically approaches the Gaussian as the spread narrows (the parameter b increases), with b acting like $1/\sigma^2$ (see Mardia, 1972). In practice, the distinction between the two distributions makes little numerical difference, especially when the spread is low; the two are are highly correlated. For example, the correlation between the distributions over the range of angular values tested in Feldman (1997) is $r > 0.999$, guaranteeing that the empirical data adduced there in support of the Gaussian transfers almost completely to the von Mises used here.

[6] These cited papers actually use the numerically very similar Gaussian distribution; see Note 5.

[7] To get a more intuitive sense of how this function behaves, consider its Taylor expansion about $\alpha = 0$, which is

the surprisal along a shape boundary[8], which closely resembles Attneave's empirical histogram plots (see also Norman et al., 2001).

The monotonic increase in suprisal with curvature does not depend on the choice of a von Mises distribution. To show this, we appeal to Chebyshev's inequality (see Feller, 1967), which provides an upper bound that applies to *all* distributions. One entailment of Chebyshev's inequality is that any probability distribution p with mean 0 obeys

$$p(x) \leq \frac{1}{z^2}, \tag{5}$$

where z is the z-score of x, that is, its value normalized by the standard deviation of the distribution. This is a rather loose bound, but one that holds regardless of the choice of distribution. In our notation, it means that for any angular distribution $p(\alpha)$ with mean 0 and standard deviation σ,

$$p(\alpha) \leq \left(\frac{\sigma}{\alpha}\right)^2, \tag{6}$$

Substituting this bound into the definition of suprisal $u(x) = -\log p(x)$, we see that the surprisal of the turning angle α is bounded below by

$$\begin{aligned} u(\alpha) &\geq -\log\left(\frac{\sigma}{\alpha}\right)^2 \\ &\geq -(\log\sigma^2 - \log[\alpha^2]) \\ &\geq \text{constant} + 2\log|\alpha|. \end{aligned}$$

In words, the surprisal increases with turning angle, at least as quickly as twice the log of its magnitude. The minus cosine increase derived above, based on the von Mises assumption, satisfies this bound. Thus regardless of the exact choice of distribution, information is bound to increase monotonically with the magnitude of the turning angle.

Curvature

Now we connect this to curvature. The curvature κ is the change in tangent direction as we move along the curve, and hence is approximated by the ratio between the change in the tangent direction (i.e., $\alpha = \Delta\phi$) and Δs:

$$\kappa \approx \frac{\alpha}{\Delta s}. \tag{7}$$

By definition, this approximation becomes exact in the limit as $\Delta s \to 0$ (i.e., as the number of sample points $n \to \infty$). Note that κ inherits its sign from α, i.e. clockwise turns are considered positive. Now rearrange terms to yield an expression for α:

$$\alpha \approx \Delta s \kappa. \tag{8}$$

We assumed above that α was distributed as a von Mises distribution with spread parameter b (Eq. 3). Because $\kappa\Delta s \approx \alpha$ this means that $\kappa\Delta s$ is distributed likewise, which in turn means that κ is distributed about 0 with spread parameter $b(\Delta s)^2$, i.e.,[9]

$$p(\kappa) \approx A' \exp[b(\Delta s)^2 \cos(\Delta s \kappa)], \tag{9}$$

where A' is again a normalizing constant (analogous to, though different from, A in Eq. 3). Plugging this into the definition of suprisal (Eq. 1), we find that the surprisal of a given value of curvature κ is

$$u(\kappa) \approx -\log A' - b(\Delta s)^2 \cos(\Delta s \kappa) \tag{10}$$

Again ignoring the additive constant (lefthand term), we see that at a given point along a curve the surprisal is proportional to the negative cosine of the product of scale and curvature,

$$u(\kappa) \propto -\cos(\Delta s \kappa) \tag{11}$$

and thus increases monotonically with curvature, exactly as Attneave proposed. (Note that the cosine function decreases monotonically from zero up through π, and thus the minus cosine increases monotonically; see Note 7.) Moreover, this expression is symmetric with respect to the sign of curvature (i.e., the surprisal is identical for κ and $-\kappa$), depending only in its magnitude—again consistent with Attneave's articulation of the claim. The details of Eqs. 10 and 11 depend on the von Mises assumption, but the Chebyshev argument above can be extended to the curvature case to yield a distribution-free bound on surprisal in terms of curvature,

$$u(\kappa) \geq \text{constant} + 2\log|\Delta s \kappa|. \tag{12}$$

The main conclusion—that suprisal increases with the magnitude of curvature—is thus guaranteed to obtain regardless of the precise choice of distribution (i.e., as long as it peaks at 0, and decreases symmetrically with increasing deviation from 0).

To be more precise, we see that in these expressions (Eq. 11 or 12), information along a contour depends on the

---

$$\begin{aligned} -b\cos(\alpha) &= -b\left[1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \ldots\right], \\ &= -b + b\alpha^2/2 + \text{higher order terms.} \end{aligned}$$

This shows in that in the neighborhood of $\alpha = 0$ (the region of psychological interest), the suprisal increases by approximately a quadratic, with the deviation from quadratic increasing as we move into the tails of the distribution. As mentioned above, the spread parameter b acts like $1/\sigma^2$, which means that $b\alpha^2/2$ likewise acts as $\alpha^2/2\sigma^2$. As it happens, this is exactly what we would have obtained had we adopted the mathematically less apt Gaussian assumption, where the exponent in the density function is $-\alpha^2/2\sigma^2$, leading to $\alpha^2/2\sigma^2$ when the minus log is taken.

[8] Code for computing surprisal based on Eq. 4 can be found at http://ruccs.rutgers.edu/∼manish/demos/curveinfo.html.

[9] Since $\frac{1}{b}$ influences the von Mises distribution in the same way that σ^2 influences the normal distribution (see Mardia, 1972), it follows that it is the quantity $\frac{1}{\sqrt{b}}$ that must be scaled by Δs; hence the new spread parameter of the von Mises would be given by $(\sqrt{b}\Delta s)^2$.

product of curvature κ and Δs. What exactly does this mean? Recall that curvature itself is not a scale-invariant quantity. When the entire figure is expanded uniformly by a given ratio (say, by inspecting it from a shorter viewing distance), all curvature values decrease by the same ratio. But because $\Delta s = L/n$, by definition Δs scales with the figure. This means that the value $\kappa\Delta s$ is scale-invariant, because whenever the figure doubles in size (say), curvature κ is halved but Δs is doubled, leaving $\kappa\Delta s$ unchanged. Another way of seeing this is to recall that the magnitude of curvature is equal to the inverse of the radius of the locally best-fitting circle, $1/R$. Hence $\kappa\Delta s = \Delta s/R = L/Rn$. But because L and R scale by the same factor, this ratio is clearly invariant to scale.[10] Δs can be thought of as the length of our "measuring stick," and the product $\kappa\Delta s$ as a measure of *scale-invariant curvature* or *normalized cuvature* (see, e.g., Hoffman & Singh, 1997; Koenderink, 1990).[11]

Hence our expression for the surprisal of curvature (Eq. 10) accords with the intuition that information along a curve is scale-invariant: it depends only on the inherent *shape* of the curve, and not on the particular viewing scale at which we happen to look at it.[12]

Closed contours

As we noted earlier, Attneave's claim refers only to the magnitude of curvature, and does not distinguish between positive and negative curvature (i.e., clockwise and counter-clockwise turning of the tangent, or equivalently, convex and concave regions). Correspondingly, our Eq. 10 is insensitive to the sign of α or κ—which followed from the fact that the distribution $p(\alpha)$ is symmetric about 0. So far, there has been no reason for it be otherwise.

However, when a visual contour is the boundary of an object—with one side of the contour assigned "figure" and the other "ground"—an asymmetry is introduced between turning one way and turning the other: one is toward figure (positive curvature), the other toward ground (negative curvature). (Our assumption that clockwise turns have positive sign means that we are travelling clockwise around the figure.) This asymmetry has been demonstrated to have clear psychological consequences.

Citing theoretical analysis and practice from art history, Koenderink and Van Doorn (1982) noted that positive curvature regions are typically perceived as having a "thing-like" character, whereas negative curvature regions are perceived as having a "glue-like" character.[13] In their seminal paper on part segmentation, Hoffman and Richards (1984) proposed that the visual system uses negative minima of curvature (points of locally highest curvature magnitude, in concave regions of a shape) to segment shapes into component parts. Thus all curvature maxima (in Attneave's sense of unsigned curvature) are not treated alike psychologically: those with negative curvature are given special status as boundaries between perceived parts, whereas equivalent ones with positive curvature are not (being perceived generally as lying on a single part).

The empirical consequences of this proposed asymmetry between positive and negative curvature (or equivalently, between convex and concave regions) have been demonstrated in a wide variety of tasks, including probe discrimination (Barenholtz & Feldman, 2003), positional judgment (Gibson, 1994; Bertamini, 2001), memory for shapes (Driver & Baylis, 1996; Braunstein, Hoffman, & Saidpour, 1989) the perception of figure and ground (Baylis & Driver, 1994; Driver & Baylis, 1996; Hoffman & Singh, 1997), amodal completion (Liu, Jacobs, & Basri, 1999), the perception of transparency (Singh & Hoffman, 1998), and visual search (Hulleman, te Winkel, & Boselie, 2000; Humphreys & Müller, 2000; Elder & Zucker, 1993).

How can the difference between positive and negative curvature be reflected in the informational analysis? Intuitively, the idea is that on a closed contour C, with the interior assumed figure, the distribution $p(\alpha)$ is "biased" so that turning in the positive-curvature direction is *more likely* than turning in the negative direction. Otherwise, the curve will not eventually close upon itself. Indeed, the geometry of curves tells exactly *how much* more likely. Over the complete circuit of the curve, the total turning angle must add up to exactly 2π (360°) of total turning angle,

$$\sum_C \alpha = 2\pi, \qquad (13)$$

which means that the expected value (mean) of the distribution $p(\alpha)$, rather than being 0 as before, must now be $2\pi/n$, where n is the number of samples taken at intervals Δs.[14]

[10] Note that this argument does *not* depend on Δs being a small or infinitesimal quantity: $L\kappa$ is a measure of scale-invariant curvature for *any* length that is tied to the scale of the figure, as all such measures are clearly proportional to one another.

[11] On 3D surfaces, one has *two* principal curvatures at each point—namely, the curvatures along the directions in which the surface curves the most and the least. Hence, it is possible to define scale-invariant notions of surface curvature by taking ratios of these quantities. Koenderink (1990), for example, defines the *shape index* in terms of the ratio $\frac{\kappa_{max}+\kappa_{min}}{\kappa_{max}-\kappa_{min}}$, a quantity that clearly remains invariant across uniform scalings. For 2D contours, however, each point has a single value of curvature associated with it—and one must thus use some measure of the scale of the figure itself to normalize the value of curvature.

[12] Our statement of this fact is of course a direct consequence of our derivation using turning angle, which is obviously scale-invariant. We emphasize it nevertheless because many discussions of shape in the literature, including casual renditions of Attneave's observation, often ignore the fact that curvature per se is scale-dependent—a tendency that requires correction in the context of any mathematical statement of the relationship between curvature and information.

[13] Koenderink and Van Doorn's analysis was developed in the context 3D-surface curvature—and more precisely, Gaussian curvature—which is somewhat more complicated than contour curvature. However, a theorem by Koenderink (1984) ensures that, for smooth surfaces, the sign of curvature of an occluding contour corresponds to the sign of Gaussian curvature on corresponding surface region. Hence their analysis transfers easily to contour curvature.

[14] Here we build on what Eq. 13 tells us about the expected mean of the distribution of turning angles on a closed curve. An alterna-

For simplicity, we assume the same von Mises form of the distribution of α as before, except with mean shifted from 0 to $2\pi/n$; that is, the entire distribution is simply translated in α-space by a small amount $2\pi/n$ in the positive direction (i.e., toward the interior of the shape):

$$p(\alpha) = A\exp[b\cos(\alpha - \frac{2\pi}{n})] \qquad (14)$$

Now substituting into the formula for suprisal as before, we get

$$u(\alpha) = -\log A - b\cos(\alpha - \frac{2\pi}{n}) \qquad (15)$$

Now we progress from angle to curvature by replacing α with its approximation $\kappa\Delta s$, and b with $b(\Delta s)^2$, yielding a formula for the suprisal as a function of curvature:

$$u(\kappa) = -\log A' - b(\Delta s)^2\cos(\kappa\Delta s - \frac{2\pi}{n}) \qquad (16)$$

Note that κ here must be interpreted in its "signed" sense with positive values assigned to the turning of the tangent towards the figure.

Here in the closed-contour case the surprisal is minimal when the tangent direction turns slightly ($2\pi/n$) inwards. Straight ($\kappa = 0$) tangents, rather than being the most expected case as before, are now slightly surprising. The key thing to observe is that points of negative curvature ($\kappa < 0$) are now *more suprising* than points of equivalent positive curvature. However much a given positive value of curvature (i.e., a turn towards the figure) is "in the tails" of the distribution—thus entailing surprise and information—the same value in the negative direction is even *more* in the tails, and hence even more suprising.

This means that negative curvature points literally carry greater information than otherwise equivalent positive-curvature points. Fig. 4 shows a plot of the information (surprisal) along a shape containing convex and concave sections of equal magnitude of curvature, illustrating the asymmetry. The magnitude of contour curvature contributes information, and negative curvature contributes more information. This picture is supported by recent empirical data showing that perceptual comparisons along the contour are generally slowed by curvature, and slowed even further by negative curvature, as compared to positive curvature of equal magnitude (Barenholtz & Feldman, 2003). The greater information content of negative curvature regions is also supported by subjects' higher sensitivity to the introduction or removal of concavity than of a comparable-sized convexity (Barenholtz, Cohen, Feldman, & Singh, 2003).

Note again that our main conclusions—that information generally increases with curvature, and is greater for concave as compared to convex turns—do not depend on the precise choice of a von Mises for the distribution (which is, though, supported by empirical data; see discussion above). Rather they follows directly from the symmetry of the distribution $p(\alpha)$ about its mean, which is required to be positive following the assumption of a closed curve. The same conclusions

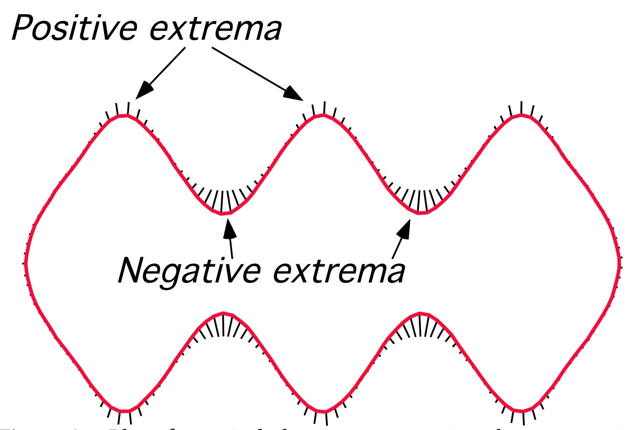

Figure 4. Plot of surprisal along a contour using the asymmetric distribution predicated on closure (Eq. 16). Note how information is greater along negatively curved (concave) portions of the contour than along positively curved (convex) portions. This shape is constructed from two mirror-image sine waves so that the indicated extrema all have identical magnitude of curvature but opposite sign (see Barenholtz & Feldman, 2003).

would have followed from any symmetric monotonically decreasing distribution, although the exact functional form of the resulting equations would be different.

It is especially interesting that no psychological assumptions about mechanisms underlying part boundary identification were necessary to derive the fact that more information is carried by negative curvature. Rather, this followed simply from the assumption of a closed curve and the implications this must have for the distribution of turning angles as the curve is traversed.

It should be noted that there exists a situation in which the contour is biased to turn *away* from the "figure" rather than toward it: namely, where a simple closed curve bounds a hole or window. In this case, the informational analysis predicts greater concentration of information in regions of the contour that are concave relative to the shaped hole, rather than concave relative to the surrounding material surface. Although this sounds counterintuitive at first, it is actually consistent with recent psychological work on the perception of holes. In particular, holes present the following perceptual anomaly: although the region surrounding the hole is clearly "figural"—in the sense of being a material surface that occludes the backdrop visible through it—the hole is nevertheless seen as a distinct perceptual entity that has its own intrinsic shape (Palmer, 1999). Thus, unlike other forms of "ground," recognition memory for the shapes of holes has

tive way to articulate the consequences of contour closure might be that instead of having n degrees of freedom (independent turning angles) in our sample, in the closed case we now have $n-1$, with that n-th being constrained to close the curve. This is a "harder" constraint, but it seems less felicitous, in part because we have no a priori reason to differentiate one of the turning angles from the other $n-1$ (as they are not of course labeled in any way). Instead we choose to employ a constraint that treats all n samples symmetrically, e.g., a softer constraint on their expected mean.

been found to be just as good as for similarly-shaped blobs (Rock, Palmer, & Hume, unpublished manuscript; cited in Palmer, 1999, p.286). From the point of view of the visual system, this means that although the surrounding surface is given a figural status as far as depth and occlusion relations are concerned, the hole is given a quasi-figural status, as far as shape analysis is concerned (Nelson & Palmer, 2001; Palmer, 1999; see also Subirana-Vilanova & Richards, 1996). Therefore, it is natural to expect that convexity relationships would be assigned relative to the hole, rather than relative to the surrounding material surface.

Applications and extensions

As Attneave suggested, specifying the distribution of information along a contour plays a key role in our understanding of how shape is mentally represented. Formalizing this observation, as we have done here, is a step towards a more rigorous understanding of shape representation. The process of formalization has, for example, already allowed us to extend Attneave's original claim—in particular, to demonstrate that regions of positive and negative curvature are not symmetric with respect to their information content, as Attneave's analysis assumed. In addition, our measure has a number of natural applications and potential extensions, some of which we mention here.

One use of a formal measure of shape information is in predicting behavioral measures pertaining to the acquisition and representation of shape information. For example, attention and eye movements may be expected to be directed towards especially informative portions of a shape. One potential difficulty, though, is that our measure only captures one kind of information, namely, information about the way a smooth contour bends (see discussion below) while attention and eye movements may be optimized to collect information more generally. A more straightforward prediction is that observers will tend to be more sensitive to *changes* in shape near highly informative regions than to those near uninformative regions, a prediction we have recently confirmed (Barenholtz et al., 2003). This connection is particularly exciting because it suggests that formal information measures might be useful more generally for predicting performance in change-detection studies—an area of enormous recent interest in the literature, but with few formal models or predictions.

Another exciting direction for development is to connect formal shape information with the underlying neural representation of shape. There is increasing interest in how neural spike trains may encode probabilities in an informationally optimal way (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1996). In the context of shape representation, recent studies involving single-cell recordings have revealed that a majority of neurons in area V4 respond preferentially to maxima of curvature magnitude along contours (Pasupathy & Connor, 2002). Moreover, distinct sub-populations of cells have been found to be selective for convex and concave extrema of contour curvature, thereby indicating that the *sign* of curvature is explicitly encoded as well. This raises the enticing possibility that there may be a direct connection between the formal information along shape contours and the underlying neural "shape code." A formal understanding of information along contours, such as that which we have presented here, is a natural first step towards the systematic investigation of such a connection.

There are number of different mathematical directions in which the information measure may be extended. One is to integrate information along the length of a contour segment, and use this integral as a measure of its *cumulative information*: $\sum_C u(\alpha)$ in the discrete case, or $\int_C u(\kappa)$ in the smooth case. This integration is particularly interesting in light of the relationship between Shannon information and complexity, widely recognized in the statistical and machine-learning literatures (Rissanen, 1989). In this literature the negative log of a probability measure is often identified with the "description length," i.e. the complexity, of a given message or pattern, because it expresses the length of the given message in an optimally efficient code (see Duda, Hart, & Stork, 2001 for an introduction). This raises the possibility that the cumulative information would serve as a psychologically realistic prediction of the descriptive complexity of a given contour—closely related to the total "bending energy" (Mumford, 1994). A formal measure of contour or shape complexity, in turn, would have numerous applications in shape perception, shape completion, and contour integration. Integrating information along an entire contour also raises the possibility that different contours could be compared in terms of their total information, i.e. their subjective shape complexity.

Caveat

Like any measure of information, ours measures information relative to the specific prior beliefs of the observer—in our case relative to the assumed prior distribution(s) over the change in tangent direction along a contour. As in Shannon's original formulation, where information is computed relative to a probability distribution over messages that is assumed known to the receiver, our measure of information is strictly predicated on particular knowledge and beliefs on the part of the observer. If these assumptions are changed, then the information measure will inevitably be changed as a result. (We explicitly consider one such possibility in the Appendix.) It is important to emphasize that the assumptions our theory attributes to the observer, as embodied in our von Mises distribution and its variants, concern *only* the way the contour is shaped locally, and do not reflect any other more global or configural properties, or indeed expectations of any other kind.

One consequence of this is that while curvature extrema maximize shape information in our sense, they do not maximize *all* kinds of information. Hence an observer seeking simply to maximize information intake would do well to turn his or her attention to a random-number generator, rather than to the curvature extrema along a shape. However while this would maximize information concerning the state of the random-number generator, it obviously would not reveal much about the shape. The point is that curvature extrema carry the most information *about the local shape of an object*, but other parts of the visual scene might well carry more information about other matters.

In this connection we should emphasize that we have *not* shown that the local shape of the object is itself, a priori, "informative" about any other parameters the observer might find of interest. A natural question to ask is whether parameters in the environment, including both shape parameters as well as others, tend to be mutually predictive (Barlow, 1994), in the sense of exhibiting high levels of mutual information (see Cover & Thomas, 1991). For example distinct tangents or turning angles along a contour might be substantially redundant in some shapes, or within parts of particular shapes. Formalizing this idea might lead to an alternative way of formalizing shape information, which we defer to future work

Conclusion

Theories of shape have often emphasized the role of curvature extrema (Richards, Dawson, & Whittington, 1986), and, in the context of perceptual part structure, negative extrema specifically (Hoffman & Richards, 1984; Hoffman & Singh, 1997; Singh, Seyranian, & Hoffman, 1999; Singh & Hoffman, 2001). It follows from our analysis that curvature extrema (in particular, positive maxima and negative minima of signed curvature) are also local maxima of information. Thus in a very concrete sense, these points carry greater information about shape than do other sections of the contour—consistent with Attneave's observation. In addition to providing mathematical justification for Attneave's claim, our analysis also extends it by demonstrating, for closed contours, a role for the sign of curvature. Whereas Attneave considered only the magnitude of curvature—treating regions of positive and negative curvature symmetrically—our analysis shows that regions of negative curvature literally carry greater information than corresponding regions of positive curvature. The psychological validity of this asymmetry is supported by empricial work on the representation of visual shape, which shows that the visual system treats regions of negative and positive curvature quite differently, and is differentially sensitive to them. Finally, our analysis also makes it clear that information attaches not to mathematical curvature per se, but rather to a normalized, scale-invariant, version of curvature ($\kappa \Delta s$ in our notation). Thus the contribution of the geometrical structure of a shape to its mental representation does not depend on scale (as curvature proper does); information is a function of "shape only" in the sense of Kendall (1977).

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.

Barenholtz, E., Cohen, E., Feldman, J., & Singh, M. (2003). Detection of change in shape: an advantage for concavities. *Cognition*, *89*(1), 1–9.

Barenholtz, E., & Feldman, J. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research*, *43*(15), 1655–1666.

Barlow, H. (1994). What is the computational goal of the neocortex? In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 1–22). Cambridge: M.I.T. Press.

Baylis, G., & Driver, J. (1994). Parallel computation of symmetry but not repetition in single visual objects. *Visual Cognition*, *1*, 377–400.

Bertamini, M. (2001). The importance of being convex: An advantage for convexity when judging position. *Perception*, *30*, 1295–1310.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Bosking, W., Zhang, Y. B. S., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex. *The Journal of Neuroscience*, *17*(6), 2112–2127.

Braunstein, M. L., Hoffman, D. D., & Saidpour, A. (1989). Parts of visual objects: an experimental test of the minima rule. *Perception*, *18*, 817-826.

Caelli, T. M., & Umansky, J. (1976). Interpolation in the visual system. *Vision Research*, *16*(10), 1055–1060.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: John Wiley.

Driver, J., & Baylis, G. (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*, 248–306.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* New York: John Wiley and Sons, Inc.

Elder, J., & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research*, *33*(7), 981–991.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 324–353.

Feldman, J. (1995). Perceptual models of small dot clusters. In I. J. Cox, P. Hansen, & B. Julesz (Eds.), *Partitioning data sets* (pp. 331–357). (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 19)

Feldman, J. (1996). Regularity vs. genericity in the perception of collinearity. *Perception*, *25*, 335–342.

Feldman, J. (1997). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835–2848.

Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, *63*(7), 1171–1182.

Feller, W. (1967). *An introduction to probability theory and its applications* (3rd ed.). New York: Wiley.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, *33*(2), 173–193.

Fisher, N. I. (1993). *Statistical analysis of circular data.* Cambridge: Cambridge University Press.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Gibson, B. (1994). Visual attention and objects: Number of objects or convexity? *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 451–470.

Gilbert, C. D. (1995). Dynamic properties of adult visual cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 73–90). Cambridge: M.I.T. Press.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*, 65–96.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*, 29–78.

Hulleman, J., te Winkel, W., & Boselie, F. (2000). Concavities as basic features in visual search: Evidence from search asymmetries. *Perception & Psychophysics*, *62*, 162–174.

Humphreys, G., & Müller, H. (2000). A search asymmetry reversed by figure-ground assignment. *Psychological Science*, *11*(3), 196–201.

Jaynes, E. T. (1973). The well-posed problem. *Foundations of physics*, *3*, 477-491.

Kendall, D. (1977). The diffusion of shape. *Advances in Applied Probability*, *9*, 428–430.

Koenderink, J., & Van Doorn, A. (1982). The shape of smooth objects and the way contours end. *Perception*, *11*, 129–137.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, *13*, 321-330.

Koenderink, J. J. (1990). *Solid shape.* Cambridge: MA: M.I.T. Press.

Liu, Z., Jacobs, D., & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research*, *39*, 4244–4257.

Mardia, K. V. (1972). *Statistics of directional data.* London: Academic Press.

Mumford, D. (1994). Elastica and computer vision. In C. L. Bajaj (Ed.), *Algebraic geometry and its applications* (pp. 491–506). New York: Springer-Verlag.

Nelson, R., & Palmer, S. E. (2001). Of holes and wholes: the perception of surrounded regions. *Perception*, *30*, 1213–1226.

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception*, *30*, 1285–1294.

Palmer, S. E. (1999). *Vision science: photons to phenomenology.* Cambridge: M.I.T. Press.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338.

Resnikoff, H. L. (1985). *The illusion of reality: topics in information science.* New York: Springer-Verlag.

Richards, W., Dawson, B., & Whittington, D. (1986). Encoding contour shape by curvature extrema. *J. Opt. Soc. Amer. A*, *3*, 1483–1491.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1996). *Spikes: exploring the neural code.* Cambridge: M.I.T. Press.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry.* Singapore: World Scientific.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Singh, M., & Hoffman, D. D. (1998). Part boundaries alter the perception of transparency. *Psychological Science*, *9*(5), 636–660.

Singh, M., & Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: segmentation and grouping in vision, advances in psychology, vol. 130* (pp. 401–459). New York: Elsevier.

Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing silhouettes: the short-cut rule. *Perception & Psychophysics*, *61*(4), 636–660.

Subirana-Vilanova, J. B., & Richards, W. A. (1996). Attentional frames, frame curves and figural boundaries: the inside/outside dilemma. *Vision Reseach*, *36*(10), 1493–1501.

Swindale, N. V. (1998). Orientation tuning curves: empirical description and estimation of parameters. *Biological Cybernetics*, *78*, 45–56.

von Mises, R. (1918). Über die "Ganzzahligkeit" der Atomgewichte und Verwandte Fragen. *Physikal Z.*, *19*, 490–500.

Appendix: Resnikoff's formulation

Resnikoff (1985) derives an information measure based on contour curvature that, he argues, mimics Attneave's proposal that information is localized in regions of extremal curvature. Resnikoff deserves credit, we feel, for placing Attneave's proposal on a mathematical footing for the first time. However, his derivation has two main problems that leave it short of providing a mathematical substantiation of Attneave's idea. First, his approach is based on the idea of gaining information by making successively finer measurements of a fixed (though unknown) quantity—which seems inappropriate when applied to the problem of measuring contour orientation at successive points along a contour. Second, the behavior of the resulting information measure comes out wrong compared to to both Attneave's claim and other psychological intuitions. In this appendix we briefly review and critique his approach.

Resnikoff's formulation is based on a general framework for quantifying the amount of information gained by successive measurements of a given parameter of fixed, but unknown, value. While Shannon's original theory assumed an observer who knows the underlying probability distribution of messages along the channel (like our shape observer, who we assume to know the distribution of turning angle along the contour), Resnikoff's theory assumes a blank-slate observer lacking this or any other prior information about the quantity in question. The question then is how successive messages (measurements) augment such an observers' knowledge.

Resnikoff's general approach is as follows. Any measurement of a parameter p has finite precision, meaning that it really consists of discovering that the parameter falls within a certain *interval* of finite non-zero size. Assume that a previous measurement has revealed has revealed p to fall within some interval (a,b) of size $|b-a|$. Now we take a second measurment and find that p falls within a smaller interval (a', b') of size $|b'-a'| < |b-a|$. How much information have we gained by taking the second measurement? Resnikoff shows that the information I (that is, really the surprisal) of the second measurement is

$$I = -\log\left(\frac{|b'-a'|}{|b-a|}\right). \tag{17}$$

This expression is very general, showing how information is transmitted via a measurement that increases precision.

Now Resnikoff relates this to curvature by applying Eq. 17 to the measurement of an angle, and specifically, the turning angle α as one moves around a smooth curve at discrete intervals Δs. Resnikoff considers that as one moves along the curve, successive measurements of the turning angle constitute successive measurements of an angle, suitable for evaluation via Eq. 17. For a given turning angle α and a given reference turning angle α_R, this gives

$$I = -\log\left(\frac{\alpha}{\alpha_R}\right), \tag{18}$$

as the information due to a given turning angle α (cf. Resnikoff's Eq. 5.2). Just as in our formulation, this can then be related directly to curvature via the relationship $\alpha = \Delta s\kappa$, to give

$$I = -\log\left(\frac{\kappa}{\kappa_R}\right) \tag{19}$$

as the expression for information as a function of curvature relative to a standard reference curvature κ_R (Resnikoff's Eq. 5.8). Resnikoff argues next that, having fixed a standard curvature κ_R, information will be extremal when curvature is extremal, exactly as Attneave proposed.

However, there are several flaws in the above argument, which we feel make Resnikoff's claim unwarranted. First, application of Eq. 17 to the case of turning angle (or curvature) seems ill-motivated. As derived and developed by Resnikoff, this equation refers to the gain in information by successive measurements of a given *fixed* quantity: that is, to changes in the state of knowledge of the observer about a fixed but unknown parameter. But turning angles at successive points along a curve do not fit this description. Turning angles have different values at different points along the curve because of the inherent geometry of the curve—the fact that it curves at different rates at different points—not because the observer has changed his or her state of knowledge about some fixed quantity. Turning angle decreases (or increases) because the curve bends, not because the observer has measured it more (or less) precisely. Hence applying Eq. 17 to turning angle does not seem valid.

Second, even accepting the validity of Resnikoff's basic set-up, the behavior of his information measure comes out wrong. As Resnikoff notes, his information measure depends always on the comparison (i.e., ratio) of two turning angles (or curvatures). Hence to evaluate the information at a particular point along a curve, one needs first a reference angle to compare it to. There are two general ways of choosing this angle, both of which Resnikoff discusses.

One is to select successive angles as one moves along the curve, comparing each turning angle to the previous one. This leads to information depending not on the turning angle, but rather on the way it (and in the smooth version, the curvature) changes as one moves along the curve. This means, extrapolating to the smooth version, that information would depend on the *derivative* of curvature with respect to arclength—not on curvature itself. This is not what Attneave proposed—and it is not, in fact, psychologically plausible. For example, it would imply that highly curved regions of a contour that were locally nearly circular would contain almost no information.

The second approach, which Resnikoff in any case favors, is to fix a reference turning angle somewhere on the curve and compare all others to it. This way, he argues, information will be extremal when turning angle, and thus curvature, is extremal with respect to this fixed standard. The problem now is that information will be extremal in the wrong way—or more precisely in one of several wrong ways depending on the choice of reference angle. Imagine that we choose a straight (zero-curvature) reference point. Now ratios of other turning angles to the reference will always be infinite (undefined, $-\log(0)$), which is clearly undesirable. So instead, select as a reference a high-curvature point. Now points with similarly-high curvature will have *low* information, while points with low curvature will have high information, exactly the opposite of Attneave's proposal. Finally, consider fixing some low-curvature point as the reference; this is Resnikoff's preference. Now regions of higher curvature will contain more information, with curvature extrema providing the most information, consistent with Attneave's proposal. However straight (zero-curvature) regions will have infinite (undefined) information, which seems qualitatively the wrong behavior.

In our formulation, in contrast with Resnikoff's, the probability of a turning angle derives not from a comparison to another one but by reference to a particular visual expectation about how smooth curves will continue, namely that they will most likely continue straight (in the open-curve case, Eq. 3). Probability is never zero and thus suprisal never infinite.

Indeed, the essential difference between our approach and Resnikoff's concerns the nature of the observer's prior assumptions about the turning angle. In Resnikoff's formulation, all the observer knows when taking a measurement is that a prior measurement revealed it to fall within a particular interval; the observer thus has has no particular expectation about *where* inside that interval the next measurement is likely to fall. This is equivalent to an assumption of *uniform probability density* over the given interval, with all values equally likely. By contrast, in our formulation, we assumed that points had been sampled from a smooth curve, so that probability density about the position of the next point was concentrated in the "forward" direction, at zero turning angle; this assumption was encapsulated in our von Mises prior. As discussed, this general form (centered at zero and monotonically decreasing away from zero—like a von Mises though not exclusively so) is supported by empirical data, and, moreoever, is related to the assumption that the points were generated by sampling a smooth curve. Hence in the context of the psychological representation of smooth contours, our non-uniform, forward-centered assumption seems justified.

However, it is well worth noting that in other contexts, something closer to Resnikoff's uniform density assumption might be appropriate. For example, if the series of vertices were generated by a fractal process, or perhaps a Brownian process with successive angles generated from a uniform density, rather than by sampling from a smooth curve, then Resnikoff's assumptions would be more apt.[15] In this case, information would follow Resnikoff's prescriptions more closely than ours. Of course, the curve resulting from such a process would little resemble the smooth contours discussed above. This raises the fascinating empirical question of whether the human visual system can "tune" its turning-angle distribution to differing environments or contexts, and if so, whether there is any way of empirically measuring the concommitant differences in the information measure. These and other questions await future research.

[15] We are grateful to Howard Resnikoff for this suggestion.