Contents lists available at ScienceDirect

# Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp



# Is perceptual space inherently non-Euclidean?

# Julian Martin Fernandez\*, Bart Farell

Institute for Sensory Research, Syracuse University, Syracuse, NY 13244, USA

#### ARTICLE INFO

## ABSTRACT

Article history: Received 5 February 2008 Received in revised form 1 October 2008 Available online 18 February 2009

Keywords: Euclidean Cue combination Bayesian Maximum likelihood Depth perception Space perception

# 1. Introduction

# 1.1. Possible perceptual spaces

In his Critique of Pure Reason, Kant (1902) argued that the truths of geometry are synthetic a priori truths. That is, three-dimensional Euclidean space is a necessary, but not tautological, presupposed form underlying all human spatial experience. Kant seems to be referring to what we today call the "intrinsic" geometry of perceptual space, rather than its "extrinsic" geometry. Extrinsic geometry refers to the relationship between the structure of the observer's perception and the actual structure of physical space. It gives the geometrical transformations necessary to map physical space onto perceptual space. Let us give an example: assume that both physical space and our internal representation of space are Euclidean, and that our percepts are distorted so that perceived objects are veridical only up to a scaling factor in depth (e.g., a circle is perceived as an ellipse). If such were the case, then the extrinsic geometry would be affine. It has long been known that many geometric relations are distorted in perception. For example, perceived distance is compressed over a large range (Gilinsky, 1951), apparent parallel allevs and equidistance allevs are not physically parallel and equidistant (Blumenfeld, 1913; Indow & Watanabe, 1984), apparent frontoparallel planes are not physically frontoparallel (Helmholtz, 1962; Ogle, 1964), and lines perceived as curved might be straight in the physical environment (Todd, Oomes, Koenderink, & Kappers, 2001).

It is often assumed that the space we perceive is Euclidean, although this idea has been challenged by many authors. Here we show that if spatial cues are combined as described by Maximum Likelihood Estimation, Bayesian, or equivalent models, as appears to be the case, then Euclidean geometry cannot describe our perceptual experience. Rather, our perceptual spatial structure would be better described as belonging to an arbitrarily curved Riemannian space.

© 2008 Elsevier Inc. All rights reserved.

Intrinsic geometry, by contrast, provides a global set of constraints by which the judgments of a given observer are formally related to one another, irrespective of their relation to the external environment. Our topic here is the structure of the intrinsic geometry of perception.<sup>1</sup> That is, we are dealing with the internal geometry of a subject's perceptual space and not at all with the relationship between perceived and actual shape. Of course, we are assuming that there is a geometry of internal perceptual space—in other words, that perception is stable and consistent enough to support such a geometry.

There are two fundamentally different ways in which intrinsic geometry could depart from being Euclidean. The first way is for it to stay within the realm of more primitive geometries, this is, geometries that do not have an internal metric structure. An instance of this is affine geometry. The second way is to keep a metric structure but to define this metric in a different way from that used in Euclidean geometry. An instance of this is Riemannian geometry.

The fact that perceived structure could be non-Euclidean<sup>2</sup> is not trivial. The interpretation of many psychophysical studies relies on the assumption of a Euclidean percept. The conclusions from any psychophysical experiment in which a variable is measured by an indirect method can potentially be affected by erroneously assuming Euclidean geometry as valid. By an "indirect method" we mean estimating the variable of interest by measuring a

<sup>\*</sup> Corresponding address: Institute for Sensory Research, Syracuse University, 621 Skytop Road, Syracuse, NY 13244, USA.

E-mail address: julian\_fernandez@isr.syr.edu (J.M. Fernandez).

<sup>0022-2496/\$ –</sup> see front matter 0 2008 Elsevier Inc. All rights reserved. doi:10.1016/j.jmp.2008.12.006

<sup>&</sup>lt;sup>1</sup> Notice that, as used in psychophysics, the words "intrinsic" and "extrinsic" have different meanings from those given in differential geometry.

<sup>&</sup>lt;sup>2</sup> We will use the term non-Euclidean to include any geometry that is not Euclidean. In some branches of geometry, the term non-Euclidean is usually restricted to meaning Riemannian geometries other than Euclidean.

different variable and linking the two variables through standard Euclidean geometry. For instance, in studies of structure from motion (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini, Caudek, & Richman, 1998) and motion–stereo depthcue combination (Domini, Caudek, & Tassinari, 2006) it is often assumed that the relationship between the perceived slant of a surface and the perceived relative depth between two points on that surface satisfies Euclidean geometry. It is unclear at this time whether and how the conclusions drawn from such studies would be affected if the Euclidean assumption were relaxed.

Siding with more primitive geometries, Gibson (1979) suggested that Euclidean metric distances in 3-dimensional space are not a primary component of an observer's perceptual experience. This hypothesis has been developed and tested by others (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini et al., 1998; Norman & Todd, 1992, 1993; Tittle, Todd, Perotti, & Norman, 1995; Todd & Bressan, 1990; Todd & Norman, 1991; Todd & Reichel, 1989). Their findings indicate that observers are quite accurate and reliable at judging an object's topological, ordinal, or affine properties and that the perception of rigid motion occurs when these properties remain invariant over time. However, accuracy is low for judgments requiring veridical perception of Euclidean metric structure, such as judgments of lengths or angles. One proposal, discussed in detail by Norman and Todd (1992), is that perceptual space can best be described as a more abstract space, in which the concept of distance, as found in Euclidean space, is not defined. Norman and Todd give the hierarchy of spatial structures that might be available to our perceptual system, from the most concrete to the most abstract (i.e., to the most primitive), as Euclidean, affine, ordinal, topological, and nominal (or categorical). This hierarchy has some resemblance to the classification of spaces given by Klein (1957), who showed that Euclidean geometry could be seen within an evolutionary sequence of more and more complex geometries: topological, affine, similarity group, Euclidean. Another, more extreme, proposal suggests that the internal structure of an object, as recovered from structure from motion, is internally inconsistent (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini et al., 1998). Thus, although observers can exhibit a conceptual understanding of Euclidean metric structure, the basis of this knowledge might be more cognitive than perceptual.

However, in all the cases described above the psychophysical evidence is also consistent with an alternative interpretation, one in which the recovered structure of the perceived object (that is, its intrinsic structure) is still essentially Euclidean but physically inaccurate. That is, the perceived shape of an object differs from the actual or simulated shape, so these experiments might reflect the properties of extrinsic rather than intrinsic geometry.

By a second alternative, mentioned above, the departure from Euclidean geometry is through a change in the metric of the space, so that the concept of distance is still defined but differently from the way in which it is defined in Euclidean geometry. This generalization of Euclidean space is Riemannian space. Riemannian geometry was first put forward in a general form by Bernhard Riemann in the nineteenth century (Petersen, 1998). It deals with a broad range of geometries whose metric properties vary from point to point, as well as two standard types of non-Euclidean geometry-spherical geometry and hyperbolic geometry-plus Euclidean geometry itself. Euclidean space, whose curvature is zero, is the simplest case of Riemannian space. The essential difference between Euclidean and Riemannian geometry is the nature of parallel lines. In Euclidean geometry, if we start with a line *l* and a point *A* not on *l*, then we can draw only one line through A that is parallel to *l*. In hyperbolic geometry, by contrast, there are infinitely many lines through A parallel to l, and in elliptic geometry, parallel lines do not exist. An example often used to make Riemannian space more intuitive is to look at what happens in 2D space. Euclidean geometry is the geometry perceived by an observer living within a flat surface (for example, a plane). A non-Euclidean geometry is the geometry perceived by an observer living within a curved surface (for example, a sphere), the surface being curved into a third spatial dimension. Riemannian geometry, as a description of perceptual space, is clearly a different alternative from those of the first of our categories, which contains more abstract (primitive) geometries—e.g., affine, topological, and similarity spaces—in which distance is not defined.

Luneburg (1947) introduced a Riemannian space of constant curvature as a description for visual space, a model that was further developed by Blank (1958, 1978). Recent experiments have shown that the assumption of a constant curvature is generally not valid (Cuijpers, Kappers, & Koenderink, 2001; Koenderink & van Doorn, 2000). For instance, Koenderink and van Doorn (2000) used classic geometry in a method that positioned the subject at the barycenter of equilateral triangles of various sizes. By remote control, the subject rotated a horizontal arrow located at one vertex of the triangle so that it appeared to point to a sphere located at another vertex. From the angle subtended by the visual direction to the arrow and the exocentric pointing direction-the veridical value was 30°-Koenderink and van Doorn obtained the perceived angle subtended by a vertex of the equilateral triangle. Then they derived the curvature of Riemannian geometry from the departure of the sum of angles of the three vertices from 180°. Koenderink and van Doorn found that the curvature changed from elliptic in near space to hyperbolic in far space, to parabolic at very large distances. Though Howard and Rogers (2002) note the possibility of measurement bias distorting subjects' settings, the results of this elegant experiment are intriguing and important.

#### 1.2. Multiple cues and the geometry of perceptual experience

Regardless of what geometry best describes our perceptual experience, shape must be obtained by combining information from the multiple cues available in the sensory input. One of the major problems in vision (and in perception in general) is to understand how the brain integrates the information provided by multiple cues. For example, by combining information from several depth cues, the visual system can estimate 3D layout with greater precision across a wider variety of viewing conditions than it could by relying on any one cue alone (Clark & Yuille, 1990). To realize this advantage, the reliability of each depth cue must be factored into the combination rule.

Several approaches have been proposed to understand how our perceptual system can solve the cue-combination problem in an effective manner. One approach to optimizing cue combination. the Maximum Likelihood Estimation (MLE) model, is statistical and uses a cue-combination rule that results in an estimator that is often unbiased and has minimum variance (for a review, see (Oruc, Maloney, & Landy, 2003)). An alternative approach is to apply Bayesian methods, in which the observer chooses an estimate that is the most probable given the image data (for a review, see (Kersten, Mamassian, & Yuille, 2004)). The two approaches give very similar estimates under many circumstances. A number of studies have tested and confirmed the quantitative predictions of the MLE model for various cues (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Landy & Kojima, 2001). Here we show that if the MLE, Bayesian, or equivalent models are valid characterizations of cue combination (as they seem to be), then Euclidean geometry cannot describe our perceptual experience. Rather, our perceptual spatial structure would be better described, at best, as belonging to an arbitrarily curved Riemannian space.

Summing up, there is ample evidence that perceptual space is not Euclidean, though there is still no consensus in the scientific community about this. As previously mentioned, many authors still treat or make the assumption that perceptual space is Euclidean. The intention of this article is to take a different approach from previous work by showing in a more explicit way why perceptual space cannot be Euclidean; the direct empirical evidence about this question is not our point. Instead, we assume first that MLE is a valid approach, based on empirical evidence, and then show that if this is so, then perceptual space is not Euclidean. In short, we will show that a Euclidean perceptual space is not compatible with MLE.

We can summarize the logic of this article as: (1) Mathematically, the shape obtained when two Euclidean shapes are combined by means of an average-an average that can be taken not just over depth values, but over any observable quantity, such as an angle-will only be Euclidean if the weights have very specific values. (2) Maximum likelihood estimated weights are determined by reliabilities (specifically, the reliability of a given cue in predicting a variable), not by geometrical constraints; one proof of this is that perceptual learning can change the weights of a given depth cue (see, e.g., Jacobs & Fine [1999]). (3) Conclusion: MLE is not Euclidean (except by chance).

# 1.3. Analysis of the inconsistency between MLE and Euclidean perceptual space

Cue combination as described by MLE and Bayesian models seems to be in close agreement with whatever procedure our perceptual system uses. They provide a rule to obtain the final estimate of a perceptual variable as the weighted average of the estimates of the individual cues. For instance, our perception of depth might combine information from motion parallax, binocular disparity, shading, texture, perspective, and so forth.

Let us assume that we have a series of variables  $x_i$  (*n* in total), each of which represents a measurable perceptual property (distance between two points, angle, etc). Individual cues, represented by the index *i* (*m* in total), each provide an estimate,  $\hat{x}_{i}$ , for each variable  $x_i$ . The estimates given by each cue are, in general, different from one another. A linear depth-cue combination model states that the observed value,  $\hat{x}_i^{obs}$ , will be:

$$\hat{x}_i^{\text{obs}} = \sum_{i=1}^m \omega_i^j \hat{x}_i^i, \tag{1}$$

where the weights  $\omega_i^j$  satisfy

m

$$\sum_{j=1}^{m} \omega_{i}^{j} = 1.$$
 (2)

(Note that MLE estimates are not necessarily minimum variancebut they are minimum variance estimates whenever the noise in the cues involved is uncorrelated (Cochran, 1963)-and not necessarily unbiased. In what follows we will assume for simplicity that these estimates are unbiased, so Eq. (1) is valid. This will not be a limitation to the generality of our results: If a different strategy other that weighted averages were used for depth-cue combination, we would still have a constraint to be satisfied, albeit this will differ from Eq. (1). The results that follow are due to the existence of a constraint in itself, regardless of its specific format, that is, whether it is given by Eq. (1) or by some other cue-combination rule. For instance, the weighted sum model is equivalent to a Bayesian model with a flat prior and flat cost function.)

In principle, there could be any number of variables and any number of cues; each additional one will increase the number of weights and thus the complexity of the problem. Thus, let us limit our analysis to the simplest case, the two-variable, two-cue scenario.<sup>3</sup> To show that nothing essential is lost by this constraint,

notice that there could be at most n - 1 independent geometric relations among the variables  $x_i$ . (A simple example of a geometric relation would be  $x_1 = 2\pi x_2$ , where  $x_1$  is the circumference of a circle and  $x_2$  is the radius.) Let us write them as:

$$\begin{aligned}
x_1 &= f_1(x_2, \dots, x_n), \\
x_2 &= f_2(x_3, \dots, x_n), \\
\dots \\
x_i &= f_i(x_{i+1}, \dots, x_n), \\
\dots \\
x_{n-2} &= f_{n-2}(x_{n-1}, x_n), \\
x_{n-1} &= f_{n-1}(x_n).
\end{aligned}$$
(3)

By substituting the equation at the bottom into the one above it, and then repeating this procedure until we reach the equation located at the top, this can be reduced to the n-1 relations:

$$x_{1} = g_{1}(x_{n}),$$

$$\dots$$

$$x_{i} = g_{i}(x_{n}),$$

$$\dots$$

$$x_{n-1} = g_{n-1}(x_{n}).$$
(4)

Here  $g_{n-1}(x_n) = f_{n-1}(x_n), g_{n-2}(x_n) = f_{n-2}(f_{n-1}(x_n), x_n)$ , and so on. This shows that the variables in Eq. (4) are related only in pairs, and thus we can restrict ourselves, without loss of generality, to the analysis of the two-variable scenario.<sup>4</sup> Let us call the two variables x and y. Then n = 2, and the set of n - 1 relations shown in Eq. (4) become the single equation

$$x = g(y). \tag{5}$$

Let us assume the best-case scenario, in which the structure recovered from each individual cue, j, is Euclidean, and thus all cues satisfy the geometric relation given by Eq. (5):

$$\hat{x}^j = g(\hat{y}^j). \tag{6}$$

The question we ask is: Does the geometric relation given by Eq. (5) still hold for the final percept? In other words, does the equation

$$\hat{x}^{\text{obs}} = g(\hat{y}^{\text{obs}}) \tag{7}$$

hold?

As mentioned, we will restrict our analysis to the simplest, twocue scenario. Call the two cues *M* and *S* (e.g., motion and stereo). Let us assume that Eq. (7) is valid. Then substituting Eq. (1) into Eq. (7) we get:

$$\omega_x^M \hat{x}^M + (1 - \omega_x^M) \hat{x}^S = g[\omega_y^M \hat{y}^M + (1 - \omega_y^M) \hat{y}^S]$$
(8)

from which we obtain:

$$\omega_x^M = \frac{g[\omega_y^M \hat{y}^M + (1 - \omega_y^M) \hat{y}^S] - \hat{x}^S}{\hat{x}^M - \hat{x}^S} = \frac{\tilde{x} - \hat{x}^S}{\hat{x}^M - \hat{x}^S}$$
(9)

where  $\tilde{x} = g[\omega_y^M \hat{y}^M + (1 - \omega_y^M) \hat{y}^S] = g(\hat{y}^{obs})$ . Now, for Eq. (7) to be true, the value of  $\omega_x^M$  given by Eq. (9) must lie in the interval [0,1] for any value of  $\omega_y^M \in [0, 1]$ . (Notice that, when  $\omega_x^M = \omega_y^M$ , Eq. (8) reduces to Jensen's inequality (see e.g., (Hardy, Littlewood, & Polya, 1934)). In such a case, the left side

 $<sup>^{3}</sup>$  The introduction of additional variables into a single relation introduces additional weights but adds nothing new to the analysis or the conclusions that

will be obtained here, which in the end are the result of having variables subject to sets of independent constraints (that is, MLE and geometric constraints).

<sup>&</sup>lt;sup>4</sup> We will not analyze the case of fewer than n-1 relations, which is not reducible to be a function of a single variable, but instead will be a function of more than one. Again, the introduction of additional variables into a single relation introduces additional weights but adds nothing new to the analysis or the conclusions that will be obtained here.



**Fig. 1.** (a) If x = g(y) is a monotonic function, then the condition expressed in Eq. (10) is fulfilled:  $\tilde{x}$  always falls between  $\hat{x}^M$  and  $\hat{x}^S$ . (b) If x = g(y) is a non-monotonic function, then the condition expressed in Eq. (10) is not valid: there are values of  $\tilde{x}$  that do not fall between  $\hat{x}^M$  and  $\hat{x}^S$ .

of Eq. (8) is usually different than the right side; they are equal only if the function *g* satisfies certain conditions). It is easy to see that if  $\omega_y^M = 0$ , then  $\omega_x^M = 0$ , and if  $\omega_y^M = 1$ , then  $\omega_x^M = 1$ . For intermediate values of  $\omega_y^M$ , this condition is equivalent to:

$$\hat{x}^{S} < \tilde{x} < \hat{x}^{M} \quad \text{if } \hat{x}^{S} < \hat{x}^{M} 
\hat{x}^{M} < \tilde{x} < \hat{x}^{S} \quad \text{if } \hat{x}^{M} < \hat{x}^{S}.$$
(10)

If x = g(y) in monotonic (see Fig. 1a), the inequality given in Eq. (10) implies that there is a single solution to Eq. (9) and thus, given  $\hat{x}^S$ ,  $\hat{x}^M$ ,  $\hat{y}^S$  and  $\hat{y}^M$ , Eq. (7) is valid only for a single pair of weights  $\omega_x^M$  and  $\omega_y^M$ . If x = g(y) is not monotonic, then there are values of  $\hat{x}^S$ ,  $\hat{x}^M$ ,  $\hat{y}^S$  and  $\hat{y}^M$  for which the inequality given in Eq. (10) is not satisfied (see Fig. 1b) and thus Eq. (7) is not true. In this latter case the percept cannot be Euclidean.

Let us concentrate then on the cases in which the Euclidean constraint could still be satisfied by having an adequate pair of weights: the monotonic case and the non-monotonic case with values of  $\hat{x}^{S}$ ,  $\hat{x}^{M}$ ,  $\hat{y}^{S}$  and  $\hat{y}^{M}$  for which the inequality given in Eq. (10) is satisfied. For these cases, Eq. (9) imposes a Euclidean constraint on perception by establishing the relationship that must hold between two parameters (the weights given to motion for two different variables, in this example). MLE as well as Bayesian and similar methods (and whatever method our visual system uses to combine cues) impose a constraint on the weights, too. These two constraints are independent: the Euclidean constraint is given by geometry, and the MLE constraint is related to the reliability of a given cue in predicting a variable. Let us look at the Cartesian space in which the two weights,  $\omega_x^M$  and  $\omega_y^M$ , represent the two coordinate axes. Cue combination constrains these two weights to take on a specific pair of values within the interval [0, 1], the values being a function of the reliabilities not only of the motion cue but also of the other cues for each of the two properties, x



**Fig. 2.** The isolated dot represents the weights prescribed by MLE or any other cue-combination process. The solid curve represents the Euclidean constraint. In general, the MLE solution and the Euclidean curve will not intersect, except by chance. If they do intersect, the weights from both would agree, that is,  $\omega_{x \text{Euclidean}}^M = \omega_{x \text{MLE}}^M$ .

and y, being estimated. Such a pair of weight values is shown as point  $(\omega_{yMLE}^{M}, \omega_{xMLE}^{M})$  in Cartesian space in Fig. 2. The Euclidean constraint from Eq. (9) will result in a curve in this space, also shown in Fig. 2. Let's assume that the actual value for  $\omega_{y}^{M}$  agrees with the value prescribed by the depth-cue combination rule. For the Euclidean constraint to be satisfied,  $\omega_{x}^{M}$  has to take a specific value on the curve ( $\omega_{xEuclidean}^{M}$  in Fig. 2). It could happen that the Euclidean constraint curve includes the point ( $\omega_{yMLE}^{M}, \omega_{xMLE}^{M}$ ) by chance; if so, then  $\omega_{xEuclidean}^{M} = \omega_{xMLE}^{M}$ . But this will not happen in general. In general, these two constraints will not give rise to the same weight values and will not be valid simultaneously. The reason is that the probability that two independent constraints result in the same weights is zero. There still remains the question of whether the departure from Euclidean (as measured by the difference between the two weights) would be large enough to be detectable and thus leading to practical consequences. The answer to this question seems to be positive. A detailed example for the case of surface slant estimation is analyzed in the Appendix.

Summing up, various studies have tested and confirmed the quantitative predictions of the MLE model for different cues, as indicated in the Introduction. The MLE constraint is independent of the Euclidean constraint given by geometry. The implication, therefore, is that perception, in general, is not Euclidean.

Implicit in this conclusion is the assumption that there is always more than one cue available. This seems like a valid assumption. There is the possibility that the visual system imposes a veto on all but one cue (for example, the most reliable one). In such an extreme case (robust estimation, see (Landy, Maloney, Johnston, & Young, 1995)) there is no cue-combination constraint to violate the Euclidean geometry constraint. However, the single-cue case is suboptimal in general, inconsistent with most laboratory evidence, and therefore not likely to represent the way our perceptual system works in daily life.

## 2. Discussion

We have shown that even if individual perceptual cues have a Euclidean structure, the final percept that emerges from cue combination will not, in general, be Euclidean. As an intuitive example, let us assume that the two variables to be estimated are the radius, r, and the circumference, c, of a circle. Let us also assume that the recovered structure from each of the two cues, M and S, is also a circle, and that the estimates of r and c obtained from each cue differ. If both M and S provide a Euclidean percept, then  $\hat{c}^M = 2\pi \hat{r}^M$  and  $\hat{c}^S = 2\pi \hat{r}^S$ . Assuming that the final percept is also be a circle, the perceived values for *c* and *r* will be (e.g., from MLE)  $\hat{c}^{obs} = \omega_c^M \hat{c}^M + (1 - \omega_c^M) \hat{c}^S$  and  $\hat{r}^{obs} = \omega_r^M \hat{r}^M + (1 - \omega_r^M) \hat{r}^S$ . In general, however, these weights will not have values that satisfy the original relationship  $c = 2\pi r$ , and so  $\hat{c}^{obs} \neq 2\pi \hat{r}^{obs}$ . The circle will inhabit a curved Riemannian space whose local curvature will depend on the particular values obtained for  $\hat{c}^{obs}$  and  $\hat{r}^{obs}$ .

Note that a percept that is non-Euclidean will not appear "strange" to the observer. An object will not pop-out just because its structure is non-Euclidean. A non-Euclidean slanted plane will still look like a Euclidean slanted plane, and a non-Euclidean circle will still look like a Euclidean circle. The only measurable effect will be that  $\pi^{obs} \neq \pi$ , but  $\pi$  in itself is not directly perceivable. However, in the case of the perception of surface slant (see Appendix), the magnitude of the departure from Euclidean structure seems to be large enough to be measured experimentally.

Our intention here was to show that perception in general cannot be Euclidean, and that "at best" it is Riemannian. This bestcase scenario is that in which (1) perception is consistent, (2) perception has an internal metric, and (3) perception based on each individual cue is Euclidean. Thus, we did not try to prove that perception is Riemannian, but only that it cannot be Euclidean. There are other scenarios, including those that might be called the "worst case". For instance, if different judgments of space were inconsistent with one another then, not only they will not be Euclidean, but they will not even be Riemannian; there will be no consistent internal geometry in such a case.

It is worth noting, however, that the MLE models do not need to, and generally do not, assume a Euclidean perceptual representation. We've shown that they couldn't validly assume it. Most of the studies found in the literature do not make any assumptions about the internal geometry of perceptual space. Some of them claim otherwise, but in practice they do not test the assumption because they usually measure only a single variable (e.g., perceived slant). They do not make any measurements that require, or test, a geometrical relationship between two different variables. Thus, most MLE results found in the literature are actually independent of the assumption of an internal Euclidean metric. Similarly, it is not necessary to assume a Euclidean metric to resolve issues of scale incompatibility between the cues. Landy et al. (1995) discussed the need to get estimates into the same units when combining cues that differ in scale. Landy et al. proposed ways of 'promoting' cues to metric scales, but cue promotion can be applied to any metrical space (that is, to any Riemannian space), whether the metric is Euclidean or not.

A few studies, however—already discussed in the previous paragraph—do assume a Euclidean relationship between variables, and thus their results should be interpreted with skepticism (but these studies are not tests of MLE, so the issue of the validity of MLE is not at stake in them).

All these cue-combination studies, and our demonstration of their implications for the geometry of internal perceptual space, seem to make an unsupported, and possibly implausible, assumption: that humans can readily estimate two or more geometric properties simultaneously. Perhaps the estimations can only be made successively, each constituting, in effect, a separate task. We assumed here the best-case scenario, in which perceptual space remains stable over time, regardless of task. If this were not true, then perceptual space would be so unstable that it would make little sense to assume that it had an internal geometry at all; the statement that perceptual space is non-Euclidean would be true without need for further argument.

Finally, it is worth noting that our demonstration is very general. It applies not only to 3D shape in vision, but also to the structure of space in general, regardless of what kind of cues lay behind its recovery—whether they come from the same modality, e.g. vision, or from different modalities, e.g., vision and touch.

#### Acknowledgments

This research was supported by NEI Grants EY015673 (J.M.F.) and EY012286 (B.F.). We thank the two anonymous reviewers for their valuable suggestions.

#### Appendix. Example: Surface slant estimation

As a concrete example, let us chose our variables to be the perceived slant of a planar surface,  $\sigma$ , and the depth difference (measured in a direction parallel to the line of sight) between two particular points on the surface,  $\Delta Z$ . We will assume the surface has a tilt of zero; in other words, iso-distant lines on the surface are horizontal. In a Euclidean space, the two variables, perceived slant and depth difference, are related as:

$$\frac{\Delta Z}{Z_0} = \frac{\Delta y}{\sigma} \tag{A.1}$$

where  $\Delta y$  is the retinal vertical separation between the points, and  $Z_0$  the distance between the observer and the surface. Eq. (A.1) is a concrete example of a geometrical relationship, as represented by Eq. (5) in the text. Let us assume that we have two depth cues, which could be, for example, texture, *T*, and binocular disparity, *D*. If the internal structure of the percepts from each of these cues is Euclidean, then they satisfy

$$\Delta \hat{z}^{T} = \frac{\Delta y}{\hat{\sigma}^{T}} \text{ and } \Delta \hat{z}^{D} = \frac{\Delta y}{\hat{\sigma}^{D}}$$
 (A.2)

were distances were normalized as  $z = Z/Z_0$ .

Based on our previous analysis, we predict that the perceived object containing the two cues T and D will depart from the Euclidean expression:

$$\Delta \hat{z}^{obs} = \frac{\Delta y}{\hat{\sigma}^{obs}} \quad \text{where} \\ \Delta \hat{z}^{obs} = \omega_z^T \Delta \hat{z}^T + (1 - \omega_z^T) \Delta \hat{z}^D \quad \text{and} \\ \hat{\sigma}^{obs} = \omega_\sigma^T \hat{\sigma}^T + (1 - \omega_\sigma^T) \hat{\sigma}^D.$$
(A.3)

The ratio  $r = \Delta \hat{z}^{obs} / (\frac{\Delta y}{\hat{\sigma}^{obs}})$  gives a measure of the departure from Euclidean perception. It is equal to 1.0 in the Euclidean case (see Eq. (A.3)), and the more *r* departs from 1.0, the more non-Euclidean the percept will become.

What we want to test here is not whether the cue-combination percept departs from being Euclidean, which has been demonstrated mathematically, but rather whether the departure could be large enough to be measurable and thus lead to practical consequences. Unfortunately, there are no experiments in the literature that simultaneously measured perceptual estimates of  $\Delta \hat{z}$  and  $\hat{\sigma}$ , for either a single-cue or a combined-cue stimulus. Thus, it is not possible at this time to test how well psychophysical data agree with our predictions. But, even if the available data are not complete, we can still obtain some order-of-magnitude estimates using the data that are available. Specifically, we can use data from the literature to obtain estimates of *r*. Using the definition of *r* and Eqs. (A.2) and (A.3), we can rewrite *r* as:

$$r = (1 + \omega_{\sigma}^{T} \delta_{2})(1 - \omega_{z}^{T} \delta_{1}) \quad \text{where}$$
  
$$\delta_{1} = \frac{\hat{\sigma}^{T} - \hat{\sigma}^{D}}{\hat{\sigma}^{T}} \quad \text{and} \quad \delta_{2} = \frac{\hat{\sigma}^{T} - \hat{\sigma}^{D}}{\hat{\sigma}^{D}}.$$
 (A.4)

For any pair of values of our choice for  $\delta_1$  and  $\delta_2$ , we can obtain reliable values for  $\omega_{\sigma}^T$  from the work of Hillis et al. (2004) who measured the reliability of slant estimation across a range of slant values for two cues, texture and stereo, and from these obtained predicted MLE weights for the cue-combination condition. They showed that the actual weights used by human observers do agree with the MLE predictions for this slant estimation task. From Fig. 5 of Hillis et al. (2004) we can see that the JNDs (Just Noticeable



**Fig. 3.** *r* as a function of  $\omega_z^T$  for the parameter values discussed in the main text extracted from Hillis et al. (2004). Any significant departure of *r* from unity implies a significant departure from a Euclidean structure.

Differences) for slant discrimination are about the same for texture and stereo cues when the slants are about 30° for texture and 0° for stereo. This happens when the viewing distance is about 100 cm. Accurate values are not very important, as we will see shortly. The important issue here is that there are many instances for which the JNDs (and thus the weights,  $\omega_{\sigma}^{T}$  and  $\omega_{\sigma}^{D}$ ) of the two cues are the same (and equal to 0.5), but the slants provided by the two cues differ. In what follows we will assume that perceived and simulated slants for individual cues are about the same, although it is known that depth from both of these cues is not always veridically perceived. This approximation will allow us to use the simulated slant values instead of the perceived ones,  $\hat{\sigma}^{T}$  and  $\hat{\sigma}^{D}$ , in Eq. (A.4).

The only parameter needed for estimating *r* that we lack is  $\omega_z^T$ . We do not have data about this weight, so Fig. 3 shows *r* as a function of  $\omega_z^T$  for the example mentioned above. Values used are  $\hat{\sigma}^T = \tan(30^\circ), \hat{\sigma}^D = \tan(0.01^\circ)$  and  $\omega_{\sigma}^T = .5$ .

Fig. 3 shows that *r* departs very considerably from 1.0 for any reasonable value of  $\omega_z^T$ : only for very small values of  $\omega_z^T$  (equivalent to a veto of the texture cue for depth difference estimation when stereo is present) is *r* close to 1.0. More dramatic departures of *r* from unity are found when using examples in which the slant difference between the texture and the stereo cues are larger, or in examples where  $\hat{\sigma}^D$  is closer to zero. With the value of *r* estimated to differ by a large amount from 1.0, departures from Euclidean structure should in fact be measurable psychophysically.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Current Biology, 14, 257–262.
- Blank, A. A. (1958). Analysis of experiments in binocular space perception. JOSA, 48, 911–925.
- Blank, A. A. (1978). Metric geometry in human binocular perception: Theory and fact. In E. L. J. Leeuwenberg, & H. F. J. M. Buffart (Eds.), Formal theories of visual perception (pp. 83–102). New York: Wiley.
- Blumenfeld, W. (1913). Unsterchunger uber die scheinbare grosse im sehraume [Studies on apparent size in visual space]. Zeitschrift fur Psychologie, 65, 241–404.

- Clark, J. J., & Yuille, A. L. (1990). Data fusion for sensory information processing systems. Kluwer.
- Cochran, W. G. (1963). Sampling techniques (2nd edition). Wiley.
- Cuijpers, R. H., Kappers, A. M. L., & Koenderink, J. J. (2001). On the role of external reference frames on visual judgments of parallelity. *Acta Psychologica*, 108, 283–302.
- Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither Euclidean nor affine. Journal of Experimental Psychology: Human Perception and Performance, 24(4), 1273–1295.
- Domini, F., & Caudek, C. (2003). 3-D structure perceived from dynamic information: A new theory. Trends in Cognitive Sciences, 10, 444–449.
- Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics*, 60, 1164–1174.
- Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. Vision Research, 46, 1707–1723.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*, 483–488.
- Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Haughton Mifflin.
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. Psychological Review, 87, 411–434.
- Hardy, G. H., Littlewood, J. E., & Polya, G. (1934). *Inequalities*. Cambridge (England): The University Press.
- Helmholtz, H. von (1962). Helmholtz's treatise on physiological optics. New York: Dover, J. P. C. Southall, Trans. (Original work published 1866).
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4, 967–992.
- Howard, I. P., & Rogers, B. J. (2002). Seeing in depth. Vol. 1. I. Porteous, Ontario, Canada.
- Indow, T., & Watanabe, T. (1984). Parallel- and distance-alleys with moving points in the horizontal plane. *Perception and Psychophysics*, 35, 144–154.
- Kant, I. (1902). Critique of pure reason. New York: American Home Library Company. Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. Annual Review of Psychology, 55, 271–304.
- Klein, F. (1957). Vorlesungen uber hohere geometrie (Lectures on higher geometry) (3rd ed.). New York: Chelsea, (Original work published 1893).
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558.
- Koenderink, J. J., & van Doorn, A. J. (2000). Direct measurement of the curvature of visual space. *Perception*, 29, 69–79.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texturedefined edges. Journal of the Optical Society of America A, 18, 2307–2320.
- Luneburg, R. K. (1947). Mathematical analysis of binocular vision. Princeton, NJ: Princeton University Press.
- Norman, J. F., & Todd, J. T. (1992). The visual perception of 3-dimensional form. In G. A. Carpenter, & S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 93–110). Cambridge, MA: MIT press.
- Norman, J. F., & Todd, J. T. (1993). The Perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, 3, 279–291.
- Ogle, K. N. (1964). Researches in binocular vision. New York: Hafner Publishing Company.
- Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. Vision Research, 43, 2451–2468.
- Petersen, P. (1998). Riemannian geometry. Berlin: Springer-Verlag.
- Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). The systematic distortion of perceived 3D structure from motion and binocular stereopsis. Journal of Experimental Psychology: Human Perception and Performance, 21, 663–678.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, 48, 419–430.
- Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, 50, 509–523.
- Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science*, 12, 191–196.
- Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, 96, 643–657.