Space Perception in Pictures

Andrea J. van Doorn[a], Johan Wagemans[b], Huib de Ridder[a], Jan J. Koenderink[b,c],

[a] Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands;
[b] Laboratory of Experimental Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium;
[c] Man Machine Interaction Group, EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

ABSTRACT

A "picture" is a flat object covered with pigments in a certain pattern. Human observers, when looking "into" a picture (photograph, painting, drawing, ... say) often report to experience a three-dimensional "pictorial space." This space is a mental entity, apparently triggered by so called pictorial cues. The latter are sub-structures of color patterns that are pre-consciously designated by the observer as "cues," and that are often considered to play a crucial role in the construction of pictorial space. In the case of the visual arts these structures are often introduced by the artist with the intention to trigger certain experiences in prospective viewers, whereas in the case of photographs the intentionality is limited to the viewer. We have explored various methods to operationalize geometrical properties, typically relative to some observer perspective. Here "perspective" is to be understood in a very general, not necessarily geometric sense, akin to Gombrich's "beholder's share". Examples include pictorial depth, either in a metrical, or a mere ordinal sense. We find that different observers tend to agree remarkably well on ordinal relations, but show dramatic differences in metrical relations.

Keywords: Vision, Picture Perception, Pictorial Space, Depth Cues, Beholder's Share

1. INTRODUCTION

When human observers look *into* pictures they are aware of a "pictorial space."[10] This pictorial space is categorically different from the space in their awareness when they look *at* the picture. In the latter case the picture and the observer are both in the same space, which may be called "the scene." The picture is part of the scene, it may be recognized as a computer screen, a photograph, drawing, or painting. In the case of a photograph it is a piece of paper covered with pigments (e.g., deposited by an ink jet printer) in a certain simultaneous arrangement. In the former case, the picture does not appear as a physical object (e.g., piece of paper in case of the photograph), but as a wormhole into another world. Pictorial space is not a proper part of the scene, it is a space that exists independently of the scene. There are no relations, for instance in terms of objects, or even the movements of objects, between the space of the scene and pictorial space. It is generally agreed on that pictorial spaces are specific to the human mind. In the animal kingdom only the primates are suspected of perhaps entertaining pictorial spaces, although opinions differ.[5]

Pictorial space is irrelevant with respect to optically guided motor actions, which goes a long way to explain why most animals appear to be oblivious to pictorial spaces. This renders pictorial spaces of much interest in the study of human perception. Pictorial perception is all about awareness and qualities, this is psychology proper.

In this paper we concentrate on space. This is not to say that the non-spatial qualities are of no interest, far from it. But space is one of the better understood aspects of visual perception, so it is of some interest to probe the spatial qualities in the pictorial domain.

A very basic property of pictorial vision that we use over and over again is that when you put a mark on a picture this mark tends to move into depth until is is "caught" by the nearest pictorial object.[10] This empirical fact can be used to design various methods to probe pictorial space. In this paper we describe a

---

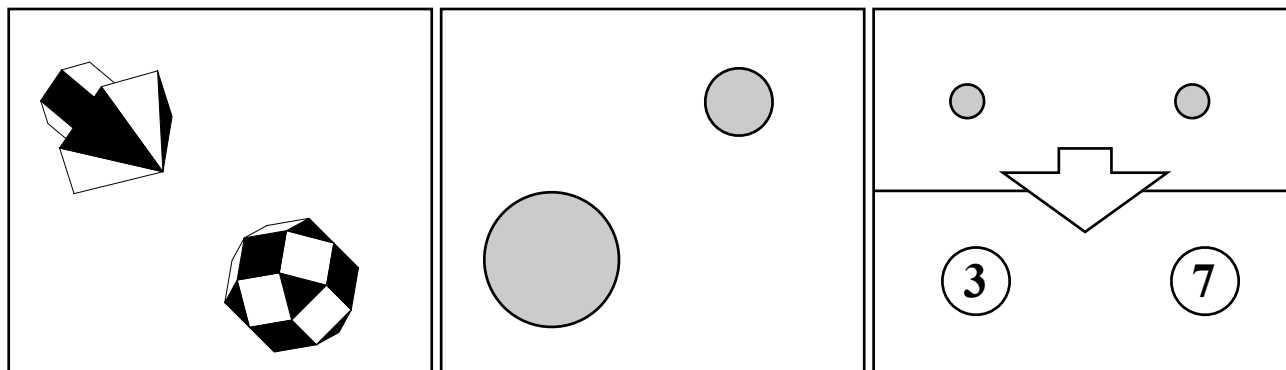Further author information: Send correspondence to a.j.vanDoorn@tudelft.nl

Figure 1. At left the pointer and target used in the pointing task (section 2.3). The observer controlled the spatial attitude of the pointer. At center the probes used in the relative size task (section 2.1). The observer controls the relative size of the objects. In the figure one has $\xi = 0.693$. At right, upper the marks used in the depth order task (section 2.2). These are simply two identical markers superimposed over the picture. At right below the appearance of the markers in the response period. In this phase the picture is absent and the markers identified by random numbers. In this case the observer would hit the "3"–key in case the left marker looked closer in the previous phase. The markers and probes shown in this figure appear superimposed over a picture in figure 3. This allows one to gain a rough notion of the structures the observers faced.

number of different methods and their use to help throw light on the issue of the existence of pictorial space. Is the hypothesis of a pictorial space useful in the description of various phenomena that can be observed in a quantitative manner, and does it allow one to describe the structure of pictorial space independent of the particular ways to probe it?

2. METHODS

It is not even that easy to define what "pictorial space" might mean. An important part of our empirical work addresses this very problem. The idea is simple enough: we attempt operational definitions of various spatial entities, then we look for correlations between the various measures, and we check whether the assumption of a "pictorial space" serves as a parsimonious description of the data. Thus the "existence" of pictorial space is understood in terms of the utility of its assumption.

In this section we describe three mutually very different operational definitions of pictorial spatial entities. We have selected examples that address the spatial configurations of mutually disjunct, more or less punctate entities. Previously we have addressed the topic of surfaces, so called pictorial reliefs. This is a fascinating topic by itself, but it is important to grasp the difference between continuous surfaces and discrete point configurations. Surfaces can be probed through local methods because they are continuously extended.[8] Discrete point configurations necessarily involve global methods though.* (See figure 1.)

2.1 Relative Size

The eye is not in pictorial space,[11] this is a basic fact that is crucial to the understanding of pictorial space. One consequence is that "distance, " that is the Euclidean distance from the eye to some fiducial point, has no equivalent in pictorial awareness. Thus "depth" has no obvious relation to "(egocentric) distance." Distance is measured in terms of some standard (meters, feet, . . .) with respect to a point, in this case the eye. "Depth" is not measured with respect to any origin. "Absolute depth" is a non-entity. In cases we talk of depth as such it is understood that depth is ambiguous and only *depth differences* truly make sense. The distance domain is the half-line (the non-negative real numbers), whereas the depth domain is the affine line (the real numbers without origin).

---

*Examples of local properties are spatial attitude or curvature as they apply to surfaces. Examples of global entities are distance or direction defined by a pair of mutually disconnected points.

Thus, in order to operationalize "depth," one cannot work with a single probe. The simplest measure involves at least a *pair* of locations. The "depth difference" for a pair of locations may perhaps be related to the distance ratio in the case of Euclidean geometry. However, it is an empirical matter whether such a psychophysical bridging hypothesis is useful. Formally, one then expects the "depth difference" to be proportional to the logarithm of the distance ratio. If this worked out in a case where "ground truth" were available, one might be induced to call the result "veridical."

How to measure depth? Since depth is a mental entity, this is not a trivial matter. One method that appears obvious is to use the well known fact that smaller visual objects look more distant. In Euclidean space angular size is simply inversely proportional with distance. In Pictorial space depth is somehow monotonically related to size: "people that look like flies are far away." It is simple enough to forge a paradigm that exploits this.

We use a method that aims at finding the depth difference between two points in pictorial space. In order to do this we superimpose two circular disks on the image (figure 1 *center*), centered at the fiducial points P_1, P_2. The diameters of the disks are $d_1 = d_0 \exp(\xi/2)$ and $d_2 = d_0 \exp(-\xi/2)$, where d_0 is some convenient fiducial size (say a few percent of the picture width), and ξ a variable on the range $\{-\infty, +\infty\}$. The geometrical mean $\sqrt{d_1 d_2} = d_0$ of the diameters is constant, whereas the logarithm of their ratio is $\xi = \log(d_1/d_2)$. These relations implement the important condition that interchanging the fiducial points P_1, P_2 corresponds to a sign change of the parameter ξ. This is obviously desirable if ξ is to be interpreted as a depth difference.

Notice that the disks have the same diameter for $\xi = 0$. The observer controls the value of the parameter ξ by means of the left-right arrow keys of the keyboard. The task is to make the disks look as to be of equal size *in pictorial space*, let the corresponding value of the parameter be denoted ξ_{12}. We then (hypothetically) treat ξ_{12} as the depth difference of the points P_1, P_2. Whether this makes sense can be put to the test by using more points. For instance, given three points P_1, P_2, P_3, we observe the values ξ_{12}, ξ_{23} and ξ_{31} (say). Notice that the order is important, that is to say $\xi_{31} = -\xi_{13}$, but that each pair can be handled in a single trial. For the hypothesis that the ξ_{ij} can be considered depth differences to make sense, one requires that $\xi_{12} + \xi_{23} + \xi_{31} = 0$. This is up to empirical verification.

More generally, given N points P_i (where $i = 1 \ldots N$), one has $N(N-1)/2$ (orderless) pairs, thus equally many observations. If the hypothesis that the ξ_{ij} (where $i, j = 1 \ldots N$, $i \neq j$) can be considered depth differences makes sense, then there should exist N numbers ξ_i (where $i = 1 \ldots N$, and $\sum_{i=1}^{N} \xi_i = 0$), such that $\xi_{ij} = \xi_i - \xi_j$, for all $i, j = 1 \ldots N$, $i \neq j$. Thus one has $N - 1$ degrees of freedom to explain $N(N - 1)/2$ independent observations. This is not trivial for $N > 3$ and it becomes a very strong test when $N \gg 2$, for then $N(N-1)/2 \gg N - 1$. In practice we estimate the ξ_i in a least squares sense, using a pseudo-inverse algorithm.

The maximum number of points that can be handled in a session of an hour with this method is about twenty. This will often be sufficient to deal with various geometrical aspects of pictorial configurations.

2.2 Depth Order

A pure depth order task can be implemented much like the relative size task. In this case the two disks can retain the same size, as the observer merely has to decide which one is the closer. The problem here is that the observer has to select the closer disk, which implies that the discs should somehow be distinguishable, thus different. This is problematic, because any difference (color, shape or size) might conceivably induce a bias. Such a bias can be accounted for by suitable balancing, but that implies each pair has to be presented twice, which clashes with the desire to work on large point sets.

In the method we implemented a trial runs as follows (figure 1 *right*): the observer is confronted with the picture, after a certain delay the two markers are superimposed, they look identical. After another delay the picture is taken away and the markers are differentiated by random labels. The observer reports the label of the marker that had appeared closer in the previous period. In practice we let the observer pace the process. This is a fast method that obviates the need for balancing. The maximum number of points that can be handled in a session of an hour is about fifty. Thus one may approach problems involving complicated configurations, the drawback being that one obtains only a rank order, no metrical data.

In the analysis we try to construct a linear depth order for N points that best accounts for the observation of $N(N - 1)/2$ observed pair orders. Notice that there exist $N!$ permutations of the numbers $1, 2, \ldots, N$, thus

finding the optimal one is a chore. We determine a good approximation by way of a voting method: the rank of a point is taken to be equal to the number of times it was judged farther.[3]

The degree to which the linear order accounts for the observations is a measure for the degree to which the pictorial space hypothesis can be accepted.

2.3 Exocentric Pointing in Pictorial Space

An entirely different method does not concern depths so much as *directions*. A direction in pictorial space has two categorically different components. One of these is the direction in the picture plane, this "tilt" component does not involve pictorial depth at all. The other component involves depth, it is the slant of the direction. The slant can be specified in terms of the ratio of depth in pictorial space and distance in the picture plane. It is a number in the range $\{-\infty, +\infty\}$. (This is the standard non-Euclidean angle definition for singly isotropic Cayley-Klein spaces.[12])

The direction from a point P_1 to a point P_2 can be measured by superimposing the picture of a target over the location of P_2 in the picture plane and a picture of a pointer over the location of P_1 in the picture plane (figure 1 *left*). Pointer and target will be experienced at their respective locations in pictorial space, that is to say, they acquire a depth, and the pointer a spatial attitude. The target is the picture of a roughly spherical object whereas the pointer is the rendering of a solid arrow, rendered in some spatial attitude. The observer is given control over this attitude and the task is to have the pointer point at the target in pictorial space. The parameters used in rendering the pointer then are converted into slant and tilt.

We find that the direction from P_1 to P_2 is not simply the reverse of the direction from P_2 to P_1, apparently the observers "point by curvilinear arcs." In the analysis we fit a unique parabolic arc at the directions, and so obtain the depth difference subtended by the points P_1, P_2. Given N points we have $N(N-1)$ ordered pairs, and thus as many observations. We try to explain these by way of a set of N depth values assigned to the points $P_1 \ldots P_2$. This involves $N-1$ degrees of freedom to fit $N(N-1)/2$ observations (each to–and–fro pointing counting as an observation).

This is a fairly slow method. The maximum number of points that can be handled in a session of an hour is about a dozen. It yields detailed and precise geometrical information though. In many cases a dozen points suffices to approach problems of interest.
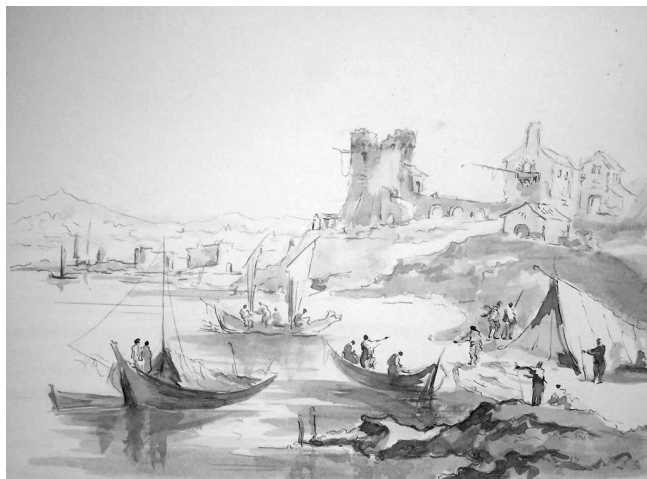

Figure 2. The picture used in the experiments. It is a wash drawing, a copy of a Capriccio by Francesco Guardi. The landscape is imaginary, apparently located in the Veneto. The construction in terms of foreground, middle ground, background, is very pronounced. Notice the use of height in the picture plane, gray tone, local contrast and edge strength as important depth cues. The size of the figures and boats serve as additional ("familiar size") depth cues, although perhaps less saliently then the aforementioned ones.

Figure 3. The picture with superimposed probes. Of course the probes are never simultaneously present, moreover they come in color whereas the picture is monochrome, and their size are increased in order to be visible in this illustration. The two smallish white points mark locations for the pairwise depth order task, The two black disks of unequal size are used in the relative size task, and the pointer and target are used in the pointing task. The marker dots (here white) are yellow, the relative size disks transparent yellow, the pointer and target red, white and pink in the actual presentations. Notice that all items are located on easily distinguished landmarks. This ensures that they will be located in pictorial space.

3. RESULTS

The methods described above allow one to put the hypothesis of the existence of a "pictorial space" to the test. This involves the value of the pictorial space assumption in each case, as well as the concordance of results over methods.

We also investigate the extent to which the obvious global depth cues are able to account for the results. Examples of such cues are height in the picture plane, local contrast range or edge strength, and so forth.

We used a single picture for all tasks (figure 3), a wash drawing by Anne-Sophie Bonno[†] after a Capriccio apparently situated in the Veneto by Francesco Guardi (1712–1793). Notice that this is an imaginary landscape, thus there is no "ground truth" of the matter. We made this choice intentionally because we are after the structure of pictorial space for which considerations of veridicality are irrelevant.

We probe this imaginary landscape with the relative size, pairwise depth order, and pointing tasks (figure 3).

3.1 The Existence of "Pictorial Space"

In the relative size method we used twenty points, this implies a hundred-and-ninety pair presentations. We find that this huge amount of observations is described very well by a set of twenty depth values with zero mean. Scatter estimated from repeated sessions allow us to test the strength of the pictorial space as a model statistically.

[†]For Anne-Sophie Bonno see `http://www.atelier-bonno.fr/`.

In the depth order task we used the same twenty points as mentioned above, but added additional ones to a total of forty-nine. This implies eleven-hundred-and-seventy-six pair presentations. We find that the linear order obtained through the voting method accounts very precisely for all pair judgments. The minor inconsistencies encountered can be contributed to scatter as estimated from repeated sessions.

In the pointing task we used five points, a subset of the points used in the other tasks. Thus we collect twenty observations. We find that a set of five depths accounts very well for the observations. Mismatches can be explained from the scatter encountered in repeated sessions.

Thus the assumption of the existence of a pictorial space indeed works well for all of the three tasks in the sense that it yields an economical description of the observations. Remaining inconsistencies are apparently due to independent random causes, as is evident from the statistical analysis of repeated sessions.
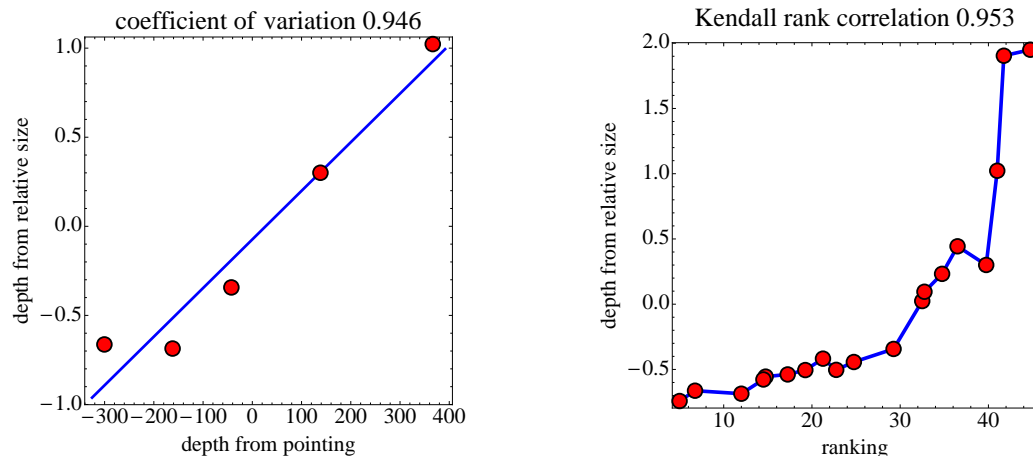

Figure 4. Some results for observer AD. At left a scatterplot between the depths from the pointing task and the relative size task. In this case there were five common locations. The coefficient of variation is $R^2 = 0.946$. At right a scatterplot between the depth order task and the relative size task. In this case there were twenty common locations. The Kendall rank correlation is 0.953. We don't need to show the relation between the rank order and the pointing task, because it was perfect. (The depth scale for the pointing is in terms of pixels. The—apparently puzzling—difference of magnitudes of depth from pointing and depth from relative size is due to the picture size, which was 1024×748 pixels. The magnitude of the ranking scale is determined by the number of points, about fifty.)

3.2 The Mutual Consistency of the Various Methods

The tasks can easily be compared, either through straight correlation, or through rank order correlation (we used the Kendall rank order measure). We find excellent agreement between the three tasks. (See figure 4.)

Various observers also come up with similar results, except for the extent of the total depth range. Whereas the depth discriminative powers are similar for all observers, the depth ranges may vary by as much as a factor of three. This is expected on theoretical grounds, and has been found in previous experiments. One expects[9] the depths for different observers to be related as

$$z_B(x, y) = \alpha_x x + \alpha_y y + \beta z_A(x, y) + \gamma,$$

where $\{x, y\}$ are the Cartesian coordinates of the picture plane, and z denotes the depth dimension for observers $\mathcal{A}$ and $\mathcal{B}$ respectively. The parameters $\{\alpha_x, \alpha_y, \beta, \gamma\}$ are necessarily idiosyncratic because left unspecified by pictorial depth cues. This relation is similar to the "bas-relief ambiguity"[1] known from machine vision. It is one aspect of Gombrich's "beholder's share."[6]

The parameter γ may safely be ignored because depth has no natural origin. The parameter β specifies depth dilations and expansions. It was first identified at the end of the nineteenth century by the German sculptor

Adolf Hildebrand.[7] The parameters $\{\alpha_x, \alpha_y\}$ describe isotropic (that is non-Euclidean) rotations. They were discovered empirically[10] and explained theoretically.[9] In machine vision they define so-called "additive planes."[1]

In the case of Guardi's *capriccio* we find that observers differ mainly in their choices of the parameter β. Some observers apparently experience more pictorial depth than others, even if their depth discriminative powers are very similar.

4. CONCLUSIONS

Apparently the question as to the existence of a pictorial space has to be answered in the affirmative. Here "existence" is understood in terms of utility, that is parsimoneous description. The assignment of depth values to points, modulo a shift—which can be accounted for by constraining the average arbitrarily to zero, suffices to explain all observations, for all tasks. The latter fact is especially encouraging because it suggests that the depths have a generic, truly *geometrical* significance, independent of their incidental operational definition. Of course this suggestion could be strengthened even further by designing additional independent ways to probe the configuration.

Such evidence for the existence of a "pictorial space" is very interesting from a psychological perspective. After all, the observer is merely confronted with a flat screen covered with gray-tones in a certain simultaneous arrangement.‡ The depth dimension is a mental entity that automatically appears in visual awareness when the observer looks "into" the picture. This three-dimensional, virtual space apparently has a tight structure that exists largely independent of the techniques used to probe it. It is a true three-dimensional geometry that has been constructed by a microgenetic process in proto-awareness.[2] It is a "presentation" in the sense that it simply happens to the observer. Of course such presentations then enter cognition and may become subject to rational thought. However, the three-dimensional geometry is present before cognition may have its go at it.

Apparently we deal with a very basic ability of the human mind, namely the ability to generate three-dimensional geometrical structures automatically, in proto–awareness, and to do so on the basis of mere pictorial cues. Neither binocular disparity, nor movement parallax, etc., are involved. It seems a priori likely that "visual space," that is the geometry of the scene in front of you when you open your eyes, might have a similar origin. However, one has to remember that most animals are unlikely to have the ability to generate pictorial presentations, even if they doubtless reveal sophisticated optically guided behavior.

ACKNOWLEDGMENTS

REFERENCES

[1] Belhumeur, P. N., Kriegman, D. and Yuille, A., "The Bas–Relief Ambiguity," International Journal of Computer Vision 35, 33–44, (1999).

[2] Brown, J. W., [Self-Embodying Mind: Process, Brain Dynamics and the Conscious Present], Barytown Ltd., Barytown NY, (2002).

[3] Condorset, Marquis de, [Essai sur lapplication de lanalyse à la probabilité des décisions rendues à la probabilité des voix], De limprimerie royale, Paris, (1785).

[4] Denis, M., "Définition du néo–traditionnisme," Art et Critique 65, 556–58, (August 23, 1890).

[5] Fagot, J., and Parron, C., "Picture perception in birds: Perspective from primatologists," Comparitive Cognition & Behavior Reviews 5, 132–135, (2010).

[6] Gombrich, E. H. J., [Art and Illusion, A Study in the Psychology of Pictorial representation], Phaidon, London, (1960).

[7] Hildebrand, A. von, [Das Problem der Form in der bildenden Kunst], (1901).

---

‡Maurice Denis:[4] "A picture, before being a war horse, a nude woman, or some anecdote, is essentially a flat surface covered by colors in a certain order."

[8] Koenderink, J. J., Doorn, A. J. van, and Kappers, A. M. L., "Surface perception in pictures," Perception & Psychophysics 52, 487–496, (1992).

[9] Koenderink, J. J., and Doorn, A. J. van, "The Structure of Visual Spaces," Journal of Mathematical Imaging and Vision 31, 171–187, (2008).

[10] Koenderink, J. J., and Doorn, A. J. van, "Pictorial space." In: [Looking into pictures: an interdisciplinary approach to pictorial space], Hecht, H., Schwartz, R, and Atherton, M. (eds.), MIT Press, Cambridge, 239–299, (2003).

[11] Wittgenstein, L. J. J., [Tractatus Logico Philosophicus], KeganPaul, London, (1922). Section 5.633.

[12] Yaglom, I. M., [A simple non–Euclidean geometry and its physical basis], Springer, New York, (1979).