Interaction of depth probes and style of depiction

Andrea J. van Doorn
Industrial Design, Delft University of Technology, Landbergstraat 15, NL-2628 CE Delft, The Netherlands;
e-mail: a.j.vandoorn@tudelft.nl

Jan J. Koenderink
EEMCS, MMI, Delft University of Technology, Mekelweg 4, NL-2628 CD Delft, The Netherlands; Laboratory of
Experimental Psychology, University of Leuven (KU Leuven), Tiensestraat 102-box 3711, BE-3000 Leuven, Belgium;
e-mail: j.j.koenderink@tudelft.nl

Mieke H. R. Leyssen
Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Tiensestraat 102-box 3711, BE-3000 Leuven,
Belgium; Interactive Information Acces, Centrum Wiskunde & Informatica, P.O. Box 94079, NL-1090 GB Amsterdam, The
Netherlands; e-mail: Mieke.Leyssen@cwi.nl

Johan Wagemans
Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Tiensestraat 102-box 3711, BE-3000 Leuven,
Belgium; e-mail: johan.wagemans@psy.kuleuven.be

Abstract. We study the effect of stylistic differences on the nature of pictorial spaces as they appear
to an observer when looking into a picture. Four pictures chosen from diverse styles of depiction
were studied by 2 different methods. Each method addresses pictorial depth but draws on a different
bouquet of depth cues. We find that the depth structures are very similar for 8 observers, apart from
an idiosyncratic depth scaling (up to a factor of 3). The differences between observers generalize
over (very different) pictures and (very different) methods. They are apparently characteristic of the
person. The differences between depths as sampled by the 2 methods depend upon the style of the
picture. This is the case for all observers except one.

1 Introduction

Representational pictorial material is very diverse. This is not just because of the subject matter, which
by no means has to be "realistic" in the sense of depicting some actual or historical event in the wid-
est sense. More importantly for this paper, pictorial material need not be in a *photographic mode*
(Gombrich, 1960). Here "photographic mode" applies equally well to imaginary scenes (say, illustra-
tions to a science fiction story) as to depictions of actual scenes. Although a watertight definition is
not possible, the photographic mode involves a commitment of the artist to at least natural perspective
and emulation of generic materials in reasonable light fields. In a more restricted definition, this would
imply linear perspective (e.g., use of a computer graphics package) and physical shading (e.g., use of
a ray tracing algorithm).

"Natural perspective" roughly coincides with Euclid's optics and concerns such issues as the
decrease of apparent size with distance, occlusions, height in the visual field, and so forth, whereas
"linear perspective" pertains to a formal projection of directions from a fixed point on a given plane.
A formal projection permits one to draw quantitative inferences from a picture. Given some additional
prior information, it may yield a veridical scale of range or distance differences up to some unknown
ambiguity. Examples of formal projections are common in architectural renderings, but very rare in the
visual arts. In this paper, we intentionally avoid such cases.

A certain form of "realism" was aimed at by the classical Western art academies (e.g., the Paris
Académie des Beaux-Arts). They offered courses that trained artists to do manually what nowadays a
computer graphics system might achieve. Realism is generally aimed for by modern computer graph-
ics. In the case of the visual arts, even in the best cases, this merely technical part was then used as a
framework for an artistic achievement. It is seen full blown in the French "*Art pompiers*" (artists such
as Jean-Léon Gérôme, William-Adolphe Bougereau, …), to mention a familiar example (Harding,

1979). This is specific to a minor corner of Western art though. "Photorealism" is used mainly as a gimmick in the arts, and even soon after the introduction of linear perspective in the Renaissance, painters intentionally deviated from it. In generic works of arts there is no such a thing as "veridical depth".

In this paper, we consider also art from other cultures and periods, cases where not even the natural perspective constraint necessarily applies. Yet, such art may be perfectly "representational" (Gombrich, 1960). Think of the Altamira or Lascaux animal, for instance.

When one considers an art work, one may either look *at* the work or *into* the work. The distinction is categorical (Koenderink, van Doorn, & Wagemans, 2011). In looking at the work, one is aware of a planar surface (we restrict ourselves here to canvasses, panels, paper, and so forth) covered with pigments in a certain arrangement. In looking into the work, one is aware of a pictorial world, not only two-fold extended, but also articulated in depth. It is this latter mode we are interested in here. Of course, this does not imply that not both aspects are of importance in the aesthetic evaluation of the work. They almost invariably are.

When looking into a work, one becomes aware of a pictorial world. Its spatial ("simultaneous presence") aspect is "pictorial space". Pictorial space may be variously articulated. In highly articulate instances, one has a feeling of "depth" at nearly every location in the picture plane. This depth is a scalar variable. Since the eye is generally not itself in pictorial space, there is no natural origin. Only depth differences are significant. Neither is there a natural unit of depth. Thus, the depth dimension is at best an affine line. In many cases, the depth dimension might be even less structured. (see Koenderink et al., 2011 for an elaboration of these notions.) "Pictorial depth" is to a large extent idiosyncratic. In the absence of a ground truth, it can only be operationally defined. Depending on the particular operationalization, observer, viewing mode, and so forth, one is bound to find different results.

We have developed a number of methods (van Doorn, Wagemans, de Ridder, & Koenderink, 2011; van Doorn, Koenderink, & Wagemans, 2011; Wagemans, van Doorn, & Koenderink, 2011a, b) that allow one to determine depth differences between locations in pictorial space that correspond to points in the picture plane. These values depend both on the picture and on the observer. They are to be considered formal elements of a first person account. The depth differences are only operationally defined, that is to say, they are created in the course of the measurement. Their values are to be considered indeterminate in the absence of an actual measurement. This distinction is key in understanding the difference between physical measurements and measurements of phenomenological experiences.

All of these methods exploit the observer's sensitivities to certain classes of depth cues (Berkeley, 1709; Gibson, 1971; Palmer, 1999; von Helmholtz, 1856). Thus, the methods are mutually distinct, and there is no reason to expect that different methods would yield similar values. The results depend (among other things) on the mix of depth cues intentionally introduced by the artist so as to evoke a certain awareness in prospective observers. They also depend upon the abilities of the observer to "read" the artist's intentions. In this study, we use only persons of generic Western cultural background as observers, whereas we vary the "style" of the works, that is to say, the bouquet of cues intentionally used by the artists. This is partly due to obvious constraints. For instance, we cannot hope to be able to use eighteenth century Japanese connoisseurs as observers in the experiment, whereas modern Japanese are intimately familiar with linear perspective through the medium of photography.

So, here are the hypotheses that we put to the test in this study. Because the various methods exploit different depth cues, it is to be expected that they will give rise to different results for instances of depiction that exploit or ignore specific depth cues. Specifically, in cases where the relative size cue is ignored, as in the case of many depictions from the non-Western cultures, we predict that depth probes based on the relative size cue will be preferentially affected. In this paper, we select methods of depth probing and styles of depiction that will likely show up such effects. Specifically, we consider styles that do or do not respect the immediate consequences of natural perspective. In doing this, we contrast paintings that lean somewhat towards photographic realism with depictions that have (at least partly) different objectives.

2 Design of the experiment

2.1 Picture material

We selected two paintings from French nineteenth century art, though not from the mainstream academic tradition, but from the contemporary avant garde. One is a painting by Caillebotte, another a painting by Pisarro (Figure 1, top row). Both are easel paintings of considerable size. Both respect
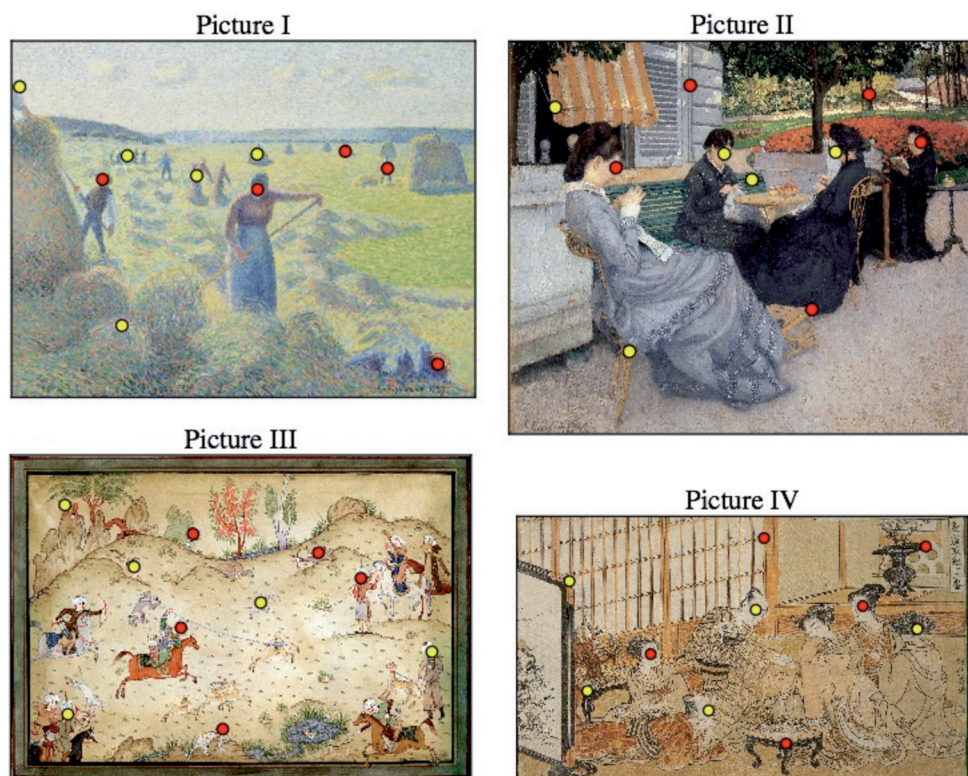
Figure 1. The four pictures used in the study. Picture I *'La Récolte des Foins, Eragny'* by Camille Pisarro, Picture II *'Portraits à la Campagne'* by Gustave Caillebotte, Picture III a miniature by Davron Toshev (Bukhara, Uzbekistan), Picture IV *'The tea ceremony'* by Isoda Koryusai. The red dots are locations used in both Methods A (pointing) and B (relative size), the yellow dots only used in Method B (relative size).

natural perspective or even a trace of linear perspective (Alberti, 1972; Gombrich, 1960; Pirenne, 1970). One consequence of this is that the size of depiction of familiar objects (say humans) depends strongly upon their intended distance from the viewer. (Just measure the size of the depicted persons in picture coordinates.) Apart from such "familiar size" cues, there are numerous additional cues to depth that are used expertly by both painters. We refer to these pictures as Picture I [*'La Récolte des Foins, Eragny'* by Camille Pisarro (1830–1903); Figure 1, top left] and Picture II [*'Portraits à la Campagne'* by Gustave Caillebotte (1848–1994); Figure 1 top right]. Notice that the Pisarro painting can hardly be said to implement formal linear perspective, whereas the Caillebotte painting has the facade in fairly precise linear perspective rendering. Even in the latter case, one hardly obtains anything like a "ground truth" without many additional assumptions. Both paintings evidently respect natural perspective.

We also selected an example from eighteenth century Japan by Koryusai, and a miniature painting by Davron (Figure 1, bottom row). Both are rather smallish pictures, printed or painted on paper in water media. In both cases, the artists to some extent ignore natural perspective, in that farther objects are not depicted noticeably smaller than closer objects of the same type. (Again, the pictures contain numerous humans.) Occlusion and height in the visual field are respected, though there is no implied eye height. The artists expertly use a number of other cues. In the case of the Japanese picture, there exists an approximate framework that is different from linear perspective but likewise offers similar formal methods of construction and measurement. The picture is constructed in an informal isometric projection and contains numerous global straight lines and nominally right angles. We refer to these pictures as Picture III [the miniature by contemporary Davron Toshev (Bukhara, Uzbekistan); Figure 1, bottom left] and Picture IV [*'The tea ceremony'* by Isoda Koryusai (1735–1790); Figure 1, bottom right].

The pictures have been selected in such a way that Pictures I (Pisarro) and III (Toshev) are comparatively "open" (both are landscapes with a far view), whereas Pictures II (Caillebotte) and IV (Koryusai) are comparatively "closed" (each a group of people engaged in some social gathering, one a garden terrace and the other a room interior). Moreover, the number and general spatial distribution of persons, vegetation, walls, furniture, and so forth, are pairwise comparable.

The pictures are presented on a computer screen [DELL U2410f monitor, 1,920 × 1,200 pixels liquid crystal display (LCD) screen], each picture filling the screen as much as possible. Thus, they are seen in similar sizes, quite unlike their actual dimensions. The widths of the pictures on the screen are 42 cm (Picture I), 38.5 cm (Picture II), 42 cm (Picture III) and 51.8 cm (Picture IV). In viewing the pictures, the observers are positioned at a fixed distance (78 cm), centered with respect to the screen. They view the screen monocularly with their dominant eye, the other eye being patched. Apart from the screen, the room was in darkness.

2.2   Participants

Eight observers participated in the experiment. Four of these observers were paid to participate and were naïve regarding the aims of the project (BC, EL, MN, RL). The remaining four were the authors (AD, JK, JW, ML).

2.3   Methods

We selected two methods that each sampled the depth difference between a pair of points in the picture. The methods are very different. One method uses probes that directly exploit the relative size cue (Wagemans et al., 2011b), whereas the other method (Wagemans et al., 2011a) does not. Thus, we expect the different styles to interact with the methods, as explained in the introduction. We refer to the methods as Method A (also referred to as "pointing") and B (also referred to as "relative size").

In Method A (pointing; see Appendix), the observer points from one point to another in pictorial space (Wagemans et al., 2011a). This involves superimposing a pointer and a target on the picture. The method does not depend explicitly on the relative size cue. We have extensive experience with this method.

In Method B (relative size; see Appendix), the observer adjusts the relative size of a pair of marks, so as to experience them as equally large in pictorial space (Wagemans et al., 2011b). Thus, this method depends crucially on the relative size cue. We also have extensive experience with this method.

In Method A (pointing), the obtained depth differences are proportional to the size of the image, in Method B (relative size) they are not. In order to render the values from the two methods comparable, we normalize the depths from Method A (pointing) by the diagonal of each image. This scaling is irrelevant for the final results.

Both methods are succinctly described in the Appendix. More extensive discussions of the methods are available elsewhere (Wagemans et al., 2011a, 2011b).

3   Experiment

3.1   Method A (pointing)

We defined five fiducial points in each picture (the red colored dots in Figure 1). These points have been chosen so as to cover the depth range. They are located so that they easily "attach" to objects in pictorial space. This latter requirement is crucial. If one puts a mark on the picture plane, it appears either as a mark on the picture plane (say, a fly speck in the sky area of a photograph) or it immerses in pictorial space and moves into depth until it attaches to the nearest (in depth) pictorial object. This happens quite naturally, as is evident from the mustaches and black teeth one often sees in poster boards depicting politicians. It is closely related to the art of painting itself, for every brush stroke (with the possible exception of the painter's signature) behaves in precisely the same way. Painting would not exist were it not for this peculiar effect, for which no scientific explanation is readily available.

Using Method A (pointing), we determine the point configuration for these five fiducial points in pictorial space. For any given picture, it turns out to be at least qualitatively similar for all observers.

Each observer repeats the measurements, doing three sessions in total. From these repeated measurements, we obtain an estimate of the significance of each observer's data. This allows us to compare observers. For any given picture, we find that although the depths for all observers correlate highly, there exist very significant differences in the overall depth ranges. There also exist very significant differences in the overall depth ranges for different pictures.

The median depth over all sessions of all observers yields a convenient overview of the results (see Figure 2). In this figure, we have plotted "ground plans" of the point configurations in depth, thus the horizontal axis is the horizontal location in the picture, whereas the vertical axis plots the depth values. We have normalized all ground plans to the same aspect ratio. A comparison of the ground plans with the locations indicated in Figure 1 will most likely agree with the intuitions of the reader. All
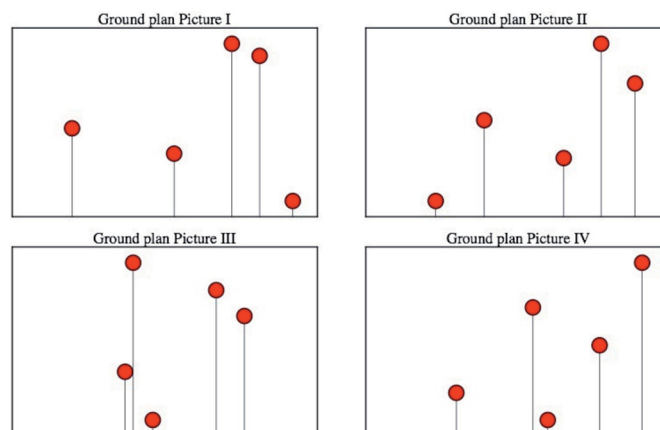
Figure 2. Ground plans of the point configurations in depth for Method A (pointing), separately for each picture (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai). The points correspond to the red points in Figure 1. The horizontal axis is the horizontal location in the picture, whereas the vertical axis plots the median depth values over all sessions of all observers. Thus, the thin lines indicate pure depth variation. (Aspect ratios have been equalized for easy reference.)

observers agreed closely on the shape (but not necessarily the magnitude) of the configuration. This is evident from the correlations of the median depth over the sessions of an observer with the overall grand median depth. The lowest coefficient of determination (R^2) over all pictures is .94, whereas typical values are .99. This indicates that all observers are aware of very similar pictorial spaces as measured by Method A (pointing).

Although the shapes of the depth configurations are indeed very similar for all observers, their depth ranges are quite different (see Figure 3). This is also evident from the slopes of the regression lines of the median depth over the sessions of an observer with the grand median depths. We find that the ratios of the largest to the smallest slopes are between 2.08 and 2.76, depending upon the picture. Whether these differences are typical of an observer may be judged by considering the profile of slope values for a given picture. When we correlate these profiles for different pairs of pictures, we find coefficients of determination (R^2) between .24 and .81 (see Table 1). Thus, observers are rather consistent over (very) different pictures. The slopes are not random, but characteristic of an observer. For a given style [Pictures I-II (Pisarro and Caillebotte) or Pictures III-IV (Toshev and Koryusai)], the correlations are quite high, whereas this is perhaps less evident when different styles are compared. For a given style, the profiles are very significantly correlated, which means that the differences are observer-specific. This holds true even when different styles are compared, although to a lesser extent.

3.2 Method B (relative size)

We defined ten fiducial points in each picture. Five of these are identical to the fiducial points used with Method A (pointing; the red colored ones in Figure 1). The additional points are shown as yellow dots in Figure 1.

The reason for the different number of fiducial points is that Method B (relative size) yields results much quicker than Method A (pointing). We do not want sessions in excess of an hour (we typically
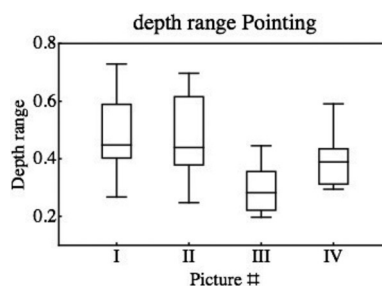

Figure 3. Depth ranges obtained with Method A (pointing) for the four pictures (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai). For each picture, we show the variation of depth ranges over all observers. The whisker box plots show the quartile ranges with the extreme values.

Table 1. Coefficients of determination (R^2) of the regressions of profiles over observers of slopes of the regression lines of the median depth over the sessions of an observer with the grand median depths separately for each picture (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai)

	I	II	III	IV
I	—	.81	.32	.6
II	.81	—	.24	.35
III	.32	.24	—	.76
IV	.6	.35	.76	—

aim at half an hour), nor do we want to distribute a task over more than a single session. From earlier experience, we have reasons to believe that the characteristics of a single observer may change over longer time spans. These are important constraints that largely determine our choices.

Using Method B (relative size), we determine the point configuration for these ten fiducial points in pictorial space. For any given picture, it turns out to be at least qualitatively similar for all observers. Each observer repeats the measurements, performing three sessions in total. From these repeated measurements, we obtain an estimate of the significance of each observer's data. This again allows us to compare observers. First conclusions are much like in the previous task. For any given picture, we find that although the depths for all observers correlate highly, there exist very significant differences in the overall depth ranges. There also exist very significant differences in the overall depth ranges for different pictures.

The median depths over all sessions of all observers are shown in Figure 4. They are plotted in the same format as Figure 2 (see legend of this figure). Again, a comparison of the ground plans with the locations indicated in Figure 1 will most likely agree with the intuitions of the reader. All observers agreed closely on the shape (but not necessarily the depth range) of the configuration. This is evident from the correlations of the median depth over the sessions of an observer with the overall grand median depth. The lowest coefficient of determination (R^2) over all pictures is .81, whereas typical values exceed .95. This indicates that all observers are aware of very similar pictorial spaces as measured with Method B (relative size).

Although the shapes of the depth configurations are indeed very similar for all observers, their depth ranges are again quite different (see Figure 5). This is evident from the slopes of the regression lines of the median depth over the sessions of an observer with the grand median depths. We find that the ratios of the largest to the smallest slopes are between 1.70 and 3.48, depending upon the picture. Whether these differences are typical of an observer may be judged by considering the profile of slope values for a given picture. When we correlate these profiles for different pairs of pictures, we find coefficients of determination (R^2) that all vary between .01 and .73 (see Table 2). For a given style [Pictures
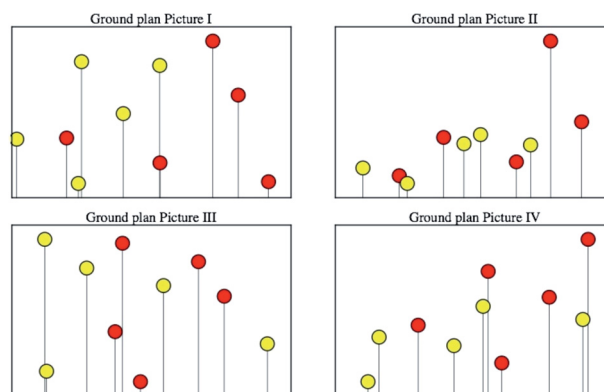

Figure 4. Ground plans of the point configurations in depth for Method B (relative size), separately for each picture (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai). The colors of the points correspond to those defined in Figure 1. The horizontal axis is the horizontal location in the picture, whereas the vertical axis plots the median depth values over all observers and all sessions. Thus, the thin lines indicate pure depth variation. (Aspect ratios have been equalized for easy reference.)
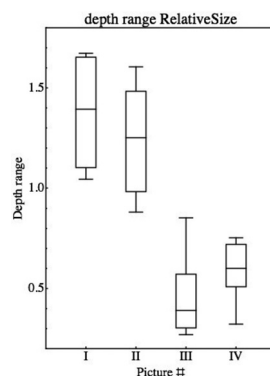
Figure 5. Depth ranges obtained with Method B (relative size) for the four pictures. For each picture (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai), we show the variation of depth ranges over all observers. The whisker box plots show the quartile ranges with the extreme values.

I-II (Pisarro and Caillebotte) or Pictures III-IV (Toshev and Koryushai)], the correlations are quite high, whereas this is not the case when different styles are compared. For a given style, the profiles are very significantly correlated, which means that the differences are observer-specific for that case.

3.3 Analysis

An important question is whether the depth configurations obtained through the two mutually very different methods are similar. This can be approached directly through a correlation of the grand median depths obtained through the two methods. Of course, we can only correlate the five points used in both Methods A (pointing) and B (relative size). We find coefficients of determination (R^2) of .91 for Picture I (Pisarro), .86 for Picture II (Caillebotte), .99 for Picture III (Toshev), and .98 for Picture IV (Koryushai). We conclude that the results of the two methods agree very well.

Another important question is whether the profiles of slopes over observers for a given picture are similar for the two methods. This can be approached through a correlation of the profiles obtained through the two methods. We find coefficients of determination (R^2) of .08 for Picture I (Pisarro), .01 for Picture II (Caillebotte), .12 for Picture III (Toshev), and .16 for Picture IV (Koryushai). We conclude that the idiosyncrasies found in either method agree hardly at all. Apparently, the differences between observers do not generalize both over pictures and over methods.

Stylistic differences might be expected when comparing Methods A (pointing) and B (relative size) for Pictures I-II (Pisarro and Caillebotte) versus Pictures III-IV (Toshev and Koryushai). This is borne out by the results. For instance, in Figure 6, we show a comparison for Pictures I (Pisarro) and III (Toshev) for observer JW. It is evident that the slopes of the regression lines both exceed one [thus, the depth range from Method B (relative size) exceeds that from Method A (pointing)] and that the slope of the regression line for Picture I (Pisarro) exceeds that for Picture III (Toshev). The fact that both slopes exceed one is in itself not remarkable (after all, the methods are quite different), but the fact that the slopes are different for the two pictures is relevant. The ratio of the slopes is a measure of the ratio of depth ranges obtained via the two methods.

The trend for observer JW (Figure 6) can be found in the results of all observers but one (JK). In Figure 7 we give whisker box plots over all observers. The ratios for Pictures I (Pisarro) and II (Caillebotte) exceed these for Pictures III (Toshev) and IV (Koryushai).

Table 2. Coefficients of determination (R^2) of the regressions of profiles over observers of slopes of the regression lines of the median depth over the sessions of an observer with the grand median depths, separately for each picture (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai).

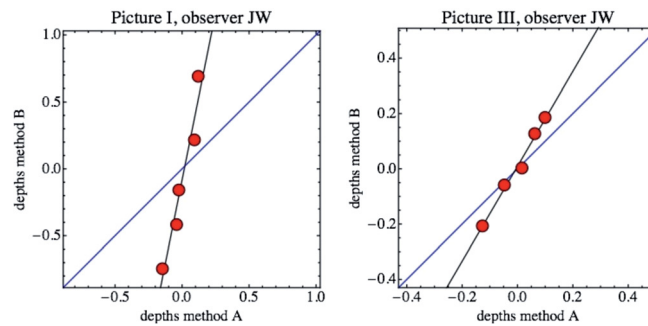|	I	II	III	IV
I	—	.73	.03	.01
II	.73	—	.07	.11
III	.03	.07	—	.67
IV	.01	.11	.67	—

Figure 6. Scatter plots of median depths obtained by Methods A (pointing) and B (relative size) for observer JW, Pictures I (Pisarro) and III (Toshev).

Thus, there is a distinction to be made for pictures that largely commit to the consequences of natural perspective [especially the dependence of apparent size on depth, Pictures I (Pisarro) and II (Caillebotte)] and those who largely ignore the issue of the dependence of apparent size on depth [Pictures III (Toshev) and IV (Koryushai)]. This was expected, as explained in the Introduction.

4 Discussion

One major result of the experiment is clear-cut: We could differentiate certain styles on the basis of the measurement of pictorial depth differences. This result is interesting in its own right, since typically the differentiation of style is left to "expert opinion" (Gombrich, 1960). It is not at all that easy to make the distinction on the basis of image characteristics. Most "obvious" measures that one might suggest actually depend upon implicit assumptions of the "expert opinion" type.

That the methods and the pictures we selected yield this result is—of course—not entirely unexpected, since we expressly selected them for this. When the interest is in a different set of stylistic elements, one would evidently select a different set of methods. Thus, we regard the present study as opening up a new realm of endeavor.

What this result vividly illustrates is the fact that "pictorial depth" is not something that is determined by the picture itself, in the sense that there would be a principled way to decide on the basis of the picture whether a set of depth assignments would be "correct". Pictorial depth as measured (otherwise it is a mere qualitative attribute someone ascribes to a picture) is necessarily operationally defined. Any method has to address a certain bouquet of "depth cues", either very generally, or rather specifically.

A most useful consequence is that methods can be focused on certain aspects of picture construction. This is the notion that we attempted to exploit in this study. We believe this concept to have numerous different applications.

We find that observers are mutually very consistent over very different pictures and very different methods in their awareness of the shape of depth configurations (Figures 2 and 4). This is encouraging
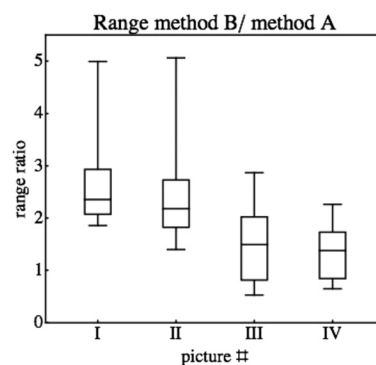

Figure 7. Ratios of depth ranges obtained with Method B (relative size) and Method A (pointing) for the four pictures (I = Pisarro, II = Caillebotte, III = Toshev, IV = Koryushai). For each picture, we show the variation of depth range ratios over all observers. The whisker box plots show the medians, interquartile ranges and the extreme values. The variation is dominated by a single outlier (observer JK).

to those who are interested in exploiting the "one picture is worth more than a thousand words" proverb. Pictures *do* convey largely unambiguous impressions, even when the style of depiction is not the preferred one in a person's cultural environment. Apparently, the visual awarenesses of our observers when looking into the pictures were very similar. Moreover, the depth configurations did hardly depend upon the method to measure them.

We found the depth *ranges* to be very idiosyncratic though. Apparently, observers experience the same pictorial space modulo some idiosyncratic depth scaling that evades our understanding. We simply have to accept that observers are different, we cannot account for it. This is essentially the message of the sculptor Hildebrand in his 1893 book *Das Problem der Form* ("On the problem of form"). Although very basic and important, this is rarely recognized in the modern literature (Hildebrand, 1893).

In summary, this type of comparative study yields very rich information about how different styles of depiction yield different types of "pictorial space" as a function of different types of measurements. There are numerous problems that might be immediately attacked by a similar approach.

References

Alberti, L. B. (1972). *On painting*. In C. Grayson (Trans.). New York, NY: Penguin Classics.

Berkeley, G.(1709). *An essay towards a new theory of vision*. Dublin, England: Aaron Rhames.

Gibson, J. J. (1971). The information available in pictures. *Viewpoints, 47*, 73–95. doi:10.2307/1572228.

Gombrich, E. H. (1960). *Art and Illusion. A study of the psychology of pictorial representation*. London, England: Phaidon Press.

Harding, J. (1979). *Artistes pompiers: French academic art in the 19th century*. New York, NY: Rizzoli.

Hildebrand, A. (1893). *Das problem der form in der bildenden kunst* [The problem of form in painting and sculpture]. Strasbourg, France: Heitz & Mündel.

Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2011). Depth. *i-Perception, 2*, 541–564. doi: 10.1068/i0438aap.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: Bradford Books/MIT Press.

Pirenne, M H. (1970). *Optics, painting and photography*. Cambridge, MA: Cambridge University Press. doi:10.1163/187847611X595891.

Pont, S., Nefs, H., van Doorn, A., Wijntjes, M., te Pas, S., de Ridder, H., et al. (2012). Depth in box spaces. *Seeing and Perceiving, 25*, 339–349. doi:10.1163/187847611X595891.

van Doorn, A. J., Wagemans, J., de Ridder, H., & Koenderink, J. J. (2011). Space perception in pictures. *Proc. SPIE, 7865*, 786519. doi:10.1117/12.882076.

van Doorn, A. J., Koenderink, J. J., & Wagemans, J. (2011). Rank order scaling of pictorial depth. *i-Perception, 2*, 724–744. doi: 10.1068/i0432aap.

von Helmholtz, H. (1856). *Handbuch der physiologischen optik* [Handbook of physiological optics]. Leipzig, Germany: Leopold Voss.

Wagemans, J. van Doorn, A. J., & Koenderink, J. J. (2011a). Measuring 3D point configurations in pictorial space. *i-Perception, 2*, 77–111. doi: 10.1068/i0420.

Wagemans, J., van Doorn, A. J., & Koenderink, J. J. (2011b). Pictorial depth probed through relative sizes. *i-Perception, 2*, 992–1013. doi: 10.1068/i0474.

Appendix

A1   Method A (pointing)

In this method (Wagemans et al., 2011a), we superimpose the pictures of a pointer and of a target on the picture (see Figure A1). The superimposed pictures are fairly small. They tend to be perceived in pictorial space, apparently moving into depth until stopped by the nearest pictorial object. The picture of the pointer is the rendering of a three-dimensional (3D) arrow. It is parameterized by slant and tilt angles. In the experiment, the observer adjusts these angles such that the pointer appears to point at the target in pictorial space.

The tilt angle is the component of the direction of the pointer in the picture plane. As expected, observers adjust the tilt such that the pointer points at the target in the picture plane. This is only to be expected, thus the tilt angle is not informative with respect to the depth structure. One finds minor systematic deviations; we have reported on such effects elsewhere. It is further ignored.

The slant angle is the component of the direction of the pointer in depth. It runs from minus 90° (pointer looks as pointing towards the viewer) to plus 90° (pointer looks like pointing purely in depth). When pointer and target are separated by a stretch L in the picture plane (say), and the slant angle is A, then we define the depth difference between pointer and target to be $L\ Tan\ A$, that is, the stretch times the tangent of the slant. We find slight differences between pointing from P to Q and pointing from Q to P. These differences are minor and idiosyncratic, though significant. We have reported on them elsewhere (Wagemans et al., 2011a). Here, we simply take the average depth difference.

After finding the depth differences between a number of points (N points say), we attempt to find the point configuration in pictorial space. Notice that there are $N\ (N-1)$ ordered point pairs, $N\ (N-1)/2$ independent values after averaging two-way pointings, as explained above. We attempt to account for these observed values in terms of N depth values. Because absolute depth remained indeterminate, we simply constrain the average depth to zero. This leaves us with $N-1$ degrees of freedom to account for $N\ (N-1)/2$ observed values. The problem is overdetermined for $N > 2$, severely so if $N >> 2$, thus we find a best solution in the least-squares sense and check consistency. We find that the observations can always be explained with an N-point configuration in pictorial space. This is crucial, the method would not be valid otherwise.

We end up with N depth values, their average being constrained to zero. As a final step, we normalize the depth values for the size of the picture. Thus, any depth difference is specified as such and so a fraction of the picture size.

A2   Method B (relative size)

In this method (Wagemans et al., 2011b), we superimpose a pair of discs on the picture (see Figure A2). The ratio of the diameters of the disks is a parameter, whereas the product of their diameters is held fixed. The disks (like the pointer and target in Method A) attach themselves to points in pictorial space. They are seen as planar disks or sometimes spheres. Observers adjust the parameter in such a way that it looks to them as if the objects were of the same size in pictorial space, the apparent size difference being ascribed to a depth difference. We take log R (the natural logarithm of the size ratio) as the depth difference. (This relation derives from very simple first principles; we have derived it in Wagemans et al., 2011b.)

Again, after finding the depth differences between a number of points (N points say), we attempt to find the point configuration in pictorial space. We attempt to account for the observed depth
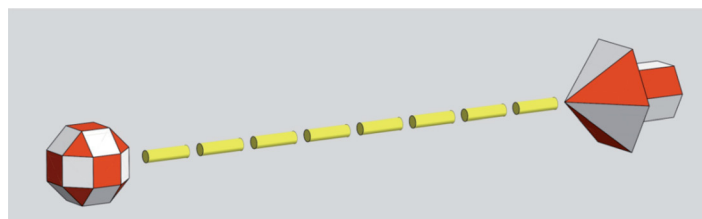

Figure A1. A "pointer" (arrow at right) and "target" (polyhedron at left) as used in the experiment. Pointer and target are superimposed over the picture. The observer may change the apparent spatial attitude of the pointer in such a way that it appears to "point to the target" [the broken yellow line in this figure indicates the pointing direction, it is (of course) not present in the actual experiment]. The spatial attitude of the pointer is defined via the 3D computer graphics; its parameters (tilt and slant angles) are used as the observer's response.
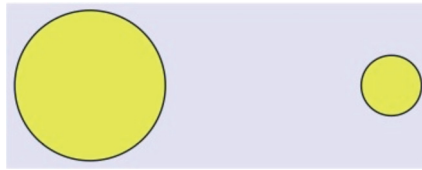
Figure A2. The relative size probe. The two disks are superposed over the picture. The geometrically mean size is held constant, the ratio of diameters being controlled by the observer. We used fairly small disks that were partially transparent. The task is to make the disks appear of equal size in pictorial space. In this figure, the disk at left would be closer than the one at right.

differences for $N(N-1)/2$ point pairs in terms of N depth values. We again constrain the average depth to zero, leaving us with $N-1$ degrees of freedom. We find that the observations can always be explained with an N-point configuration in pictorial space. This is crucial, the method would not be valid otherwise. We end up with N depth values, their average being constrained to zero.

This method has been described in detail elsewhere (Wagemans et al., 2011b).

A3   Heuristic on depth range ratios

In this paper, we compare ratios of depth ranges. It is indeed necessary to do so, because depth ranges per se are irrelevant. It is not possible to predict the depth range ratio from first principles. In this appendix, we present a heuristic. It has the additional advantage that it illustrates a number of potential misunderstandings.

That there is no principled method to predict the depth range ratio is clear from the following observations. In the pointing method, the depth differences (and thus the depth ranges) are proportional to the size of the picture. In the relative size method, the differences (and thus the depth ranges) are independent of the size of the picture. *Ergo*, the ratio of depth ranges is dependent on the size of the picture, it is not some fixed number. One way to "correct" for this is to scale the depth differences from the pointing method by the size of the picture. This is what we do in the paper. It is essentially arbitrary. However, in the context of the paper, this is not problematic at all because (in the final instance) we only compare depth range ratios.

It remains an interesting issue what a typical depth range ratio might be (of course, taking picture size into account). One possible heuristic is to consider a result we published a short time ago (Pont et al., 2012). Observers view a rendering of a cube in "one point perspective" and are given the task to adjust the perspective so as to look "like a cube". They are placed at diverse distances from the screen, and the picture comes in many sizes. From an understanding of perspective, one expects the observers to adjust the perspective in a predictable manner. Perhaps surprisingly, they don't. Human observers prefer a standard "template" view over the "veridical" one. Deviations from the template look not cubical, but like long corridors or thin slabs.
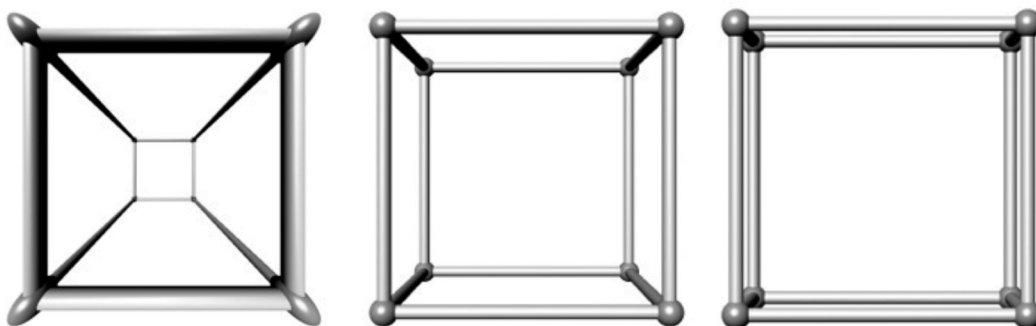

Figure A3. A cube seen in "one-point perspective" (implying that there is only a single finite vanishing point, it is the centre of each subfigure). At left, the viewing distance is very short, thus this is a wide-angle picture. (Notice the perspective distortions of the spheres that mark the corners.) As measured in the picture plane, the back pane is much smaller than the front pane. In reality they are equally large, though this is hard to "see". This looks more like a deep corridor than like a cube. At right, the viewing distance is very large, thus this is a telephoto picture. Front and back panes are almost equally large in the picture plane. This looks more like a flattish slab than like a cube. At the centre is a view that most observers accept as a typical "picture of a cube". It looks like a cube from any viewing distance.
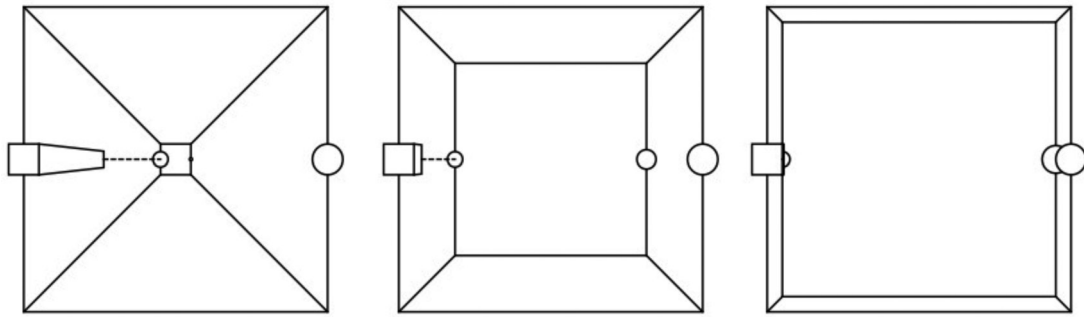
Figure A4. Here the one points cube perspectives have been augmented with depth probes. At the left side of each subfigure, a small cube acting as "pointer" in front and a "target" at the back. Pointer and target implement Method A. The pointing is "correct" by construction. At the right side of each subfigure, two disks, one in front and the other at the back. The two disks implement Method B. The relative size ratio is "correct" by construction. These constructions are a standard application of linear perspective as conventionally taught to artists and technical draftsmen.

In the one-point perspective (Figure A3), we may easily construct a pointer with target and a pair of relative size probes (Figure A4). This takes nothing but the conventional perspective constructions, using vantage and distance points. From these synthetic probes, we may find the depth difference between pointer and target (divided by the edge length) and the depth difference from the relative sizes. We simply apply the same calculations as used in the assessment of settings using Method A (pointing) or B (relative size). Thus, we obtain a ratio of depth ranges. It is 0.183 for the preferred perspective. For the long corridor (an extreme perspective), it is 5.76, and for the thin slab 0.029.

The ratio is higher when the perspective is stretched into depth, as when using a wide angle lens on a camera, whereas it is lower when the perspective is squashed, as when using a telephoto lens on a camera. Thus, the heuristic leads one to predict that the ratio of depth ranges will be higher for Pictures I (Pisarro) and II (Caillebotte) than for Pictures III (Toshev) and IV (Koryusai), which we indeed find to be the case.

Andrea van Doorn (1948) studied physics, mathematics and chemistry at Utrecht University, where she did her master's in 1971. She did her Ph.D. (at Utrecht) in 1984. She is presently at Delft University of Technology, Department of Industrial Design. Current research interests are various topics in vision, communication by gestures, and soundscapes.

Jan Koenderink (1943) studied physics, mathematics and astronomy at Utrecht University, where he graduated in 1972. From the late 1970's he held a chair "The Physics of Man" at Utrecht University till his retirement in 2008. He presently is Research Fellow at Delft University of Technology, and guest professor at the University of Leuven. He is a member of the Dutch Royal Society of Arts and Sciences, and received a honorific doctorate in medicine from Leuven University. Current interests include the mathematics and psychophysics of space and form in vision, including applications in art and design.

Mieke Leyssen (1987) has an M.Sc. in experimental psychology and artificial intelligence, both obtained at the University of Leuven. As part of her training she did an internship in Stephen Palmer's Visual Perception and Aesthetics Lab at the University of California, Berkeley. Currently, she is working as a Ph.D. student in the Interactive Information Access research group at the Centrum Wiskunde & Informatica (CWI) in Amsterdam. Her research interests include human-computer interaction, user-interface design, cognitive science and visual perception.

Johan Wagemans (1963) has a BA in psychology and philosophy, an MSc and a PhD in experimental psychology, all from the University of Leuven, where he is currently a full professor. Current research interests are mainly in so-called mid-level vision (perceptual grouping, figure-ground organization, depth and shape perception) but stretching out to low-level vision (contrast detection and discrimination) and high-level vision (object recognition and categorization), including applications in autism, arts, and sports (see www. gestaltrevision.be).