The viewpoint-dependency of veridicality: psychophysics and modelling

Bert Willems, Johan Wagemans *

Laboratory of Experimental Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium

Received 29 November 1999; received in revised form 9 May 2000

Abstract

Human observers were shown projected angles, embedded in solid cross-like figures and were asked whether these projected angles could be the projection of an orthogonal angle in 3-D space (i.e. whether the two legs of the cross were orthogonal to each other). We found that performance depended on the viewpoint at which the angle was viewed: Both slant (i.e. the angle between the normal of the target angle relative to the plane of projection) and roll (i.e. the rotation around the normal of the target angle) had a systematic effect on the proportion of errors when observers were shown non-orthogonal angles. With orthogonal angles, however, this effect was absent (i.e. very low error rate with no systematic effect of slant and roll). Instead of assuming a viewpoint-dependent bias towards orthogonality, a computational analysis of the task, using a Bayesian approach, and a computer simulation showed that the viewpoint-dependency can be modelled by a fixed set of biases in order to constrain the set of possible scenes that could give rise to the projection. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Viewpoint-dependency; Veridicality; Psychophysics and modelling

1. Introduction

Visual perception is often conceived as a process of inference (e.g. Knill & Richards, 1996): Properties of the 3-D world have to be inferred from the projection of these properties onto our 2-D retina. However, these inferences are not deductively valid: Because of the loss of information about the depth of each point in the world (i.e. the distance between this point and its projection onto our retina), there is an infinite set of 3-D worlds that can give rise to a particular projection. However, despite the non-deductive nature of this process of inference, we have a clear percept of a stable 3-D world. For example, objects presented to us have a clear 3-D structure that appears not to be changing when changing the viewpoint from which the objects are seen.

In order to explain why we can make judgements about the 3-D structure of objects presented to us, it is assumed that this process of perceptual inference is constrained by a priori assumptions about the world we are looking at (Knill & Kersten, 1991). For example, the orientation of a plane surface can be inferred using the assumption that the texture on the surface is homogeneously distributed and has no particular orientation bias (as is mostly the case with plane surfaces presented to us; Knill, 1998). Another example is the use of the rectangularity constraint for inferring surface orientation: the slant of three planes can be inferred assuming that the three planes meet at orthogonal angles to each other (Attneave & Frost, 1969).

However, using these a priori assumptions in order to constrain the set of possible 3-D interpretations leads to veridical perception only when the statistical structure of the constraints is the same as the statistical structure of the corresponding properties of our environment (i.e. the constraints are ecological). When this is not the case, these inferences can lead to non-veridical perception: The 3-D interpretation inferred from the retinal projection does not correspond to the 3-D scene that gave rise to this projection.

This is clearly illustrated by some perceptual illusions. For example, a precisely constructed set of planes

---

* Corresponding author. Tel.: +32-16-325969; fax: +32-16-326099.

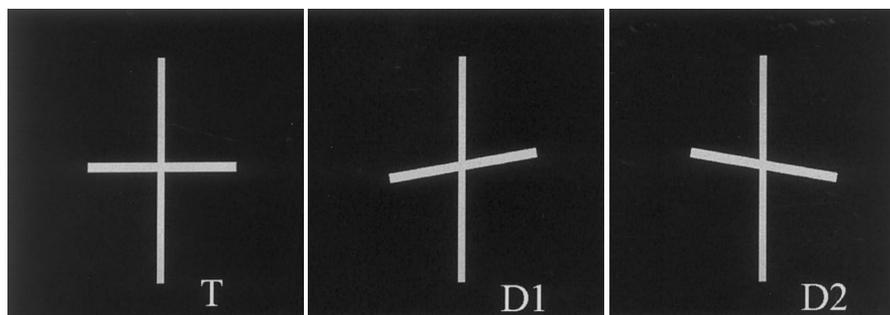*E-mail address:* johan.wagemans@psy.kuleuven.ac.be (J. Wagemans).

Fig. 1. The three objects used in the experiment. The three objects are shown in the fronto-parallel plane and the 3-D deviation from orthogonality for the two distractors equals $-10°$ ($\Phi_{3D} = 80°$) and $10°$ ($\Phi_{3D} = 100°$) for D1 and D2 respectively (with the upper-right angle as reference angle).

can result in non-veridical perception of the relations between these planes, as is the case with the Ames room (Ittelson, 1952; see also Perkins & Cooper, 1980 and Griffiths & Zaidi, 2000, for some related cases). Another example is the situation of a precisely calculated movement (an object moving in anti-phase with the moving observer) that can give rise to non-veridical perception (i.e. the object was not seen as moving, but was seen as smaller and closer, due to the induced motion-parallax; Wallis & Bülthoff, 1998). The notion 'precisely constructed' already indicates the low probability of these events in our 3-D world. This is probably the reason why these constraints are imposed on the process of perceptual inference: Our perception remains veridical as long as the things in the world behave and are constructed according to the statistical structure of our process of perceptual inference.

In this study, we present some evidence that in some cases the veridicality of our percept depends on the viewpoint at which the objects are presented to us. Observers were presented solid cross-like figures with the two legs orthogonal to each other or not (see Fig. 1). We found that in some cases distractors (cross-like figures with the two legs not orthogonal to each other) were seen as being orthogonal (i.e. non-veridical perception). This result can easily be explained by assuming a bias toward the rectangular interpretation. However, the probability of a distractor being seen as an orthogonal cross-like figure was also dependent on viewpoint, suggesting that the strength of the biases, needed to constrain the set of possible interpretations, is dependent on viewpoint: for some orientations the bias toward the rectangular interpretation is stronger than for others. However, a precise analysis of the computational demands of this task shows that this seemingly complex performance can be explained by a simple computational model (Bayesian inference), in which the constraints, used by our visual system, do not depend on viewpoint.

In Section 2, we present psychophysical evidence that the veridicality (whether non-orthogonal angles are interpreted as being orthogonal or not) of our percept

does depend on viewpoint. Then, we formally describe the task faced by our observers and present a computational model that is able to simulate these results (Section 3). Finally, the psychophysical results will be discussed in light of the results obtained from the computational analysis (Section 4).

2. Psychophysical data

2.1. Methods

2.1.1. Participants

Four members of the laboratory of Experimental Psychology participated in this experiment, three graduate students who were naive about the purpose and the methodological details of the experiment and the second author. All of them had normal or corrected-to-normal vision. The small number of participants was justified by the large number of trials and by the fact that all individuals produced highly similar data.

2.1.2. Apparatus

The experiment was carried out on a PC with a Pentium 133 MHz processor. Stimuli were displayed on an SVGA computer screen with 1024×768 spatial resolution and 75 Hz temporal resolution. Responses were given by pressing the Z-key (orthogonal angle) or the M-key (oblique angle) on a QWERTY keyboard.

2.1.3. Stimuli

Using the graphics software of 3D Studio (Autodesk, Inc., 1993), we created an orthogonal target-cross (T) and two oblique distractor-crosses (D1 and D2); see Fig. 1 for a front-view of the three objects. If we take the upper-right quadrant of the crosses as reference angle (denoted by the symbol Φ_{3D}), the deviation from orthogonality for the two distractors was $-10°$ and $+10°$ for D1 and D2, respectively ($\Phi_{3D}(D1) = 80°$, $\Phi_{3D}(T) = 90°$ and $\Phi_{3D}(D2) = 100°$). Restricting our attention to only this quadrant in the following analyses does not invalidate the results because every angle can
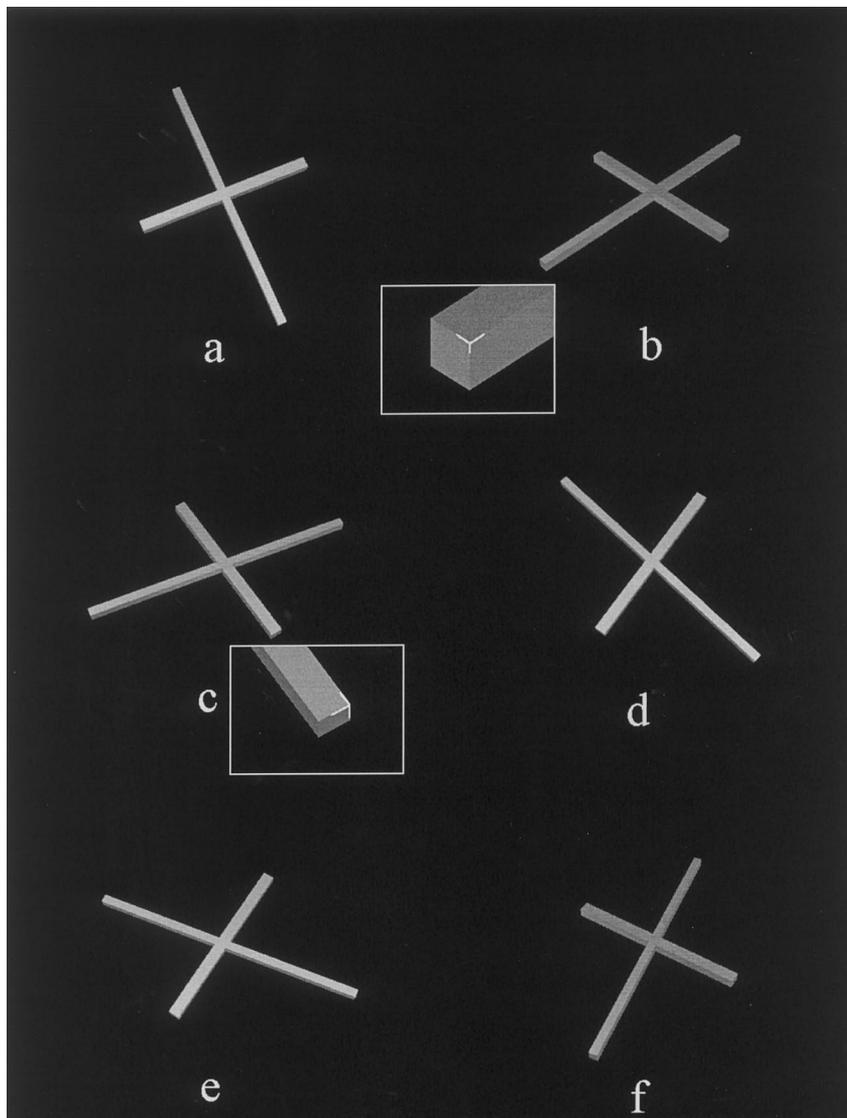
Fig. 2. Some example images used in the experiment. (a) T with $\rho = 22.5°$ and $\sigma = 20°$. (b) T with $\rho = -45°$ and $\sigma = 50°$. (c) D1 with $\rho = -67.5°$ and $\sigma = 30°$. (d) D1 with $\rho = 45°$ and $\sigma = 20°$. (e) D2 with $\rho = 67.5°$ and $\sigma = 20°$. (f) D2 with $\rho = -22.5°$ and $\sigma = 40°$. The images in the insets illustrate the information (Y- and arrow-junctions at the end of the legs) that could be used to derive the orientation of the target-angle in 3-D space.

be expressed in terms of the other one. Note also that the two distractors can be regarded as two images of the same 3-D object, differing from each other by a 180° rotation around the Y-axis (i.e. a reflection). However, for reasons that will become clear later, we decided not to regard it as one distractor but as two different distractors.

All crosses were copper coloured and were $18 \times 12°$ of visual angle when upright in the frontoparallel plane.

Perspective projection was used to render the crosses (the virtual camera was always directed towards the centre of the cross) but the maximal difference of the depth coordinates within the crosses was sufficiently small relative to the focal distance to regard the deformation as negligible, in line with the assumption of orthographic projection used throughout the paper. The horizontal leg crossed the vertical leg in the middle and the thickness of the two legs was such that the Y- and arrow-junctions at the end of the legs were clearly visible (see insets in Fig. 2). The reason we used solid figures instead of isolated angles is that these Y- and arrow-junctions provide useful depth information. Because our participants knew that the angles composing the junctions were orthogonal in 3-D space (i.e. participants knew that the only uncertainty was at the angles in the middle of the cross), the orientation of the angle relative to the plane of projection could be calculated in principle (Attneave & Frost, 1969).

An additional source of information about the actual position of the cross with respect to the observer could

be the length of the two legs of the cross (certainly after a number of trials). Based on these two sources of information, it was possible (in principle) for our observers to calculate the exact orientation of the objects. If we would not provide our observers with these additional sources of information (i.e. a projected angle in isolation with variable length of legs), the problem of inferring the 3-D structure of the projected angle would be ill-posed and could not be solved in principle (see Section 3).

The position of the cross in space was varied by first rotating the cross in the picture plane (i.e. manipulating the roll of the cross; rotation about the Z-axis where the picture plane is the XY-plane) and then slanting it in depth (rotation about the X-axis). No additional rotation about the Z-axis (i.e. the tilt) was applied. Each object (T, D1, and D2) was depicted at all 35 positions in 3-D space by combining seven levels of roll ($-67.5°$, $-45°$, $-22.5°$, $0°$, $22.5°$, $45°$, $67.5°$ with positive values for counterclockwise roll) with five levels of slant (20, 30, 40, 50, 60°), leading to 105 trials. Such a basic block of trials was repeated 20 times for each subject resulting in a total number of 2100 trials per subject. Some example images are shown in Fig. 2.

2.1.4. Procedure

The experiment was controlled by a program developed in Superlab Pro (Cedrus, Corp., 1997). A trial consisted of a black screen for 1000 ms followed by an image of a cross, depicted from a particular position in space. This remained on the screen until the subjects responded. The screen was then made black again for 1000 ms until the next trial began (leading to an effective intertrial interval of 2 s). The subjects' task was to tell for each image on the screen whether it depicted a target object or a distractor object by pressing the Z-key or M-key, respectively. They were asked to respond quickly (i.e. within a second) to minimise the use of non-perceptual reasoning strategies and they were told that the proportion of target- and distractor-images was not necessarily equal.

The experiment was performed individually in a dark room, with participants seated at one meter from the monitor, with their head on a chin rest. By showing some examples, it was made clear to the subjects that distractor objects were derived from the target object only by changing the angle between the two legs while preserving the coplanarity between them (e.g. the horizontal leg could not bend forward). None of the subjects had difficulties understanding the task. A block of 105 trials required about 8–10 min. Subjects could take several blocks within one session, interrupted by short breaks. Sessions were distributed over several days.

2.2. Results

For each participant, we calculated the error rate (i.e. the proportion of errors calculated over the 20 measurements) for each type of object (T, D1, and D2), for each value of roll and for each value of slant. For the target-cross, this is called the *miss rate* while for the two distractors this is called the *false-alarm rate*. These data were entered in an analysis of variance (ANOVA). All variables are within-subject variables. All effects were statistically reliable at the 0.0001 level (see Table 1). The most interesting of these effects, the differential effects of slant and roll on error rate, depending on object type, are shown in Fig. 3a and b, respectively.

Table 1

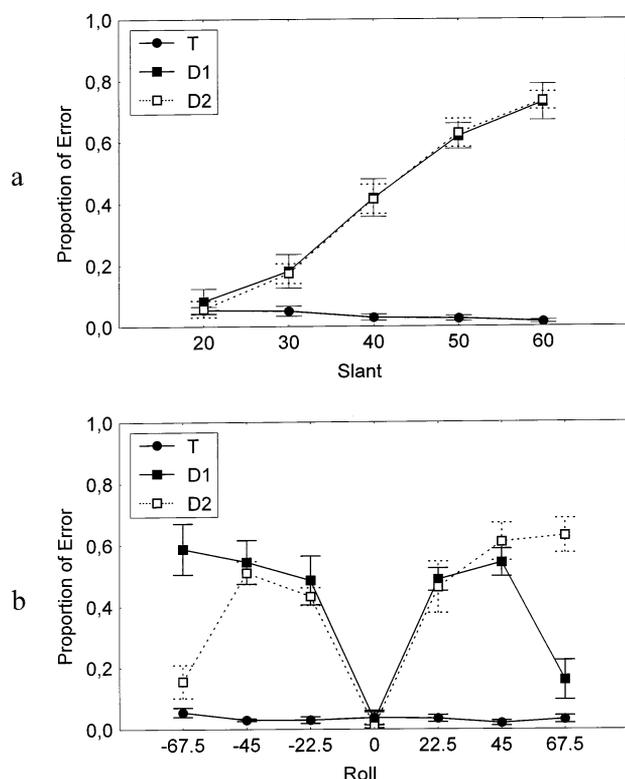| Effect	F	df	MSe	P
Rotation angle	37.05	6, 18	0.02	0.000001
Slant angle	137.01	4, 12	0.02	0.000001
Object type	65.51	2, 6	0.10	0.0001
Rotation $\times$ slant	12.14	24, 72	0.01	0.000001
Rotation $\times$ object	18.69	12, 36	0.03	0.000001
Slant $\times$ object	69.97	8, 24	0.01	0.000001
Rotation $\times$ slant $\times$ object	4.86	48, 144	0.01	0.000001

Fig. 3. Results of the experiment (averages and standard errors calculated over the four different subjects). (a) Effect of slant-angle on the proportion of error for each type of object (averaged over the different roll-values). (b) Effect of roll on the proportion of error for each type of object (averaged over the different slant-values).

The basic pattern of results can be summarised as follows:

- When confronted with a *target* ($\Phi_{3D} = 90°$), perception is veridical (i.e. very low miss rates). Together with the high false alarm rates, this suggests that our participants had a bias toward the orthogonal interpretation.
- When confronted with *distractor 1* ($\Phi_{3D} = 80°$), the veridicality of our participants' performance deteriorates with the slant of the angle (resulting in high false-alarm rates; see Fig. 3a) and is systematically worse for negative values of roll than for positive values of roll (see Fig. 3b).
- When confronted with *distractor 2* ($\Phi_{3D} = 80°$), the veridicality of our participants' performance deteriorates again with the slant of the angle (resulting in high false-alarm rates; see Fig. 3a) and is systematically worse for positive values of roll than for negative values of roll (see Fig. 3b).

2.3. Discussion

Observers were asked whether projected angles (embedded in solid figures) were orthogonal angles in 3-D space or not. We found that whether our observers' perceptions were veridical or not (i.e. whether orthogonal angles were seen as being orthogonal and non-orthogonal angles were seen as being non-orthogonal) depended heavily on the viewpoint at which the angles were seen. For some orientations, non-orthogonal angles were more likely to be interpreted as being orthogonal than for others, leading to the suggestion that the biases, needed to constrain the set of possible interpretations, are dependent on the viewpoint at which the angles are viewed.

However, this is a most unlikely situation. Because it can be assumed that through our evolution and personal development, the constraints used by our visual system reflect the statistical structure of our natural environment, these constraints are very unlikely to be dependent on viewpoint because the visual environment does not change with changing viewpoint.

Fortunately, this kind of explanation can be avoided when we consider the computational aspects of this process of perceptual inference in more detail. In the following sections, we will analyse the statistical structure of this process of inference by means of a simple Bayesian observer (with a fixed bias toward the rectangular interpretation) that is able to simulate this viewpoint-dependent performance without relying on a changing set of constraints with viewpoint. The results of this computational analysis show two important facts. First, the information provided by the images changes for the different orientations of the object relative to the plane of projection. This gives rise to the viewpoint-dependency in the performance of our ob-

servers. Second, this change in information is a natural consequence of the statistical structure of the task and not of a change in the representations involved.

3. Computational analysis of the task

The task faced by our observers is to infer from a given projected angle whether this angle is orthogonal in 3-D space or not. This is no trivial task because the projected angles (i.e. Φ_{2D}) are not only dependent on the actual shape of this angle in 3-D space (i.e. Φ_{3D}) but also on the orientation of this angle relative to the plane of projection. Let us define each scene as a vector **S**, composed of

- the 3-D angle, Φ_{3D}, ranging from 0 to π,
- the slant σ of the angle, that is, the angle between the line of sight and the normal of the 3-D angle (i.e. the direction orthogonal to both legs of the 3-D angle), ranging from 0 to $\pi/2$ and
- the roll ρ of the 3-D angle, that is, the rotation around the normal of the 3-D angle, ranging from $-\pi$ to π.

Now, when a particular scene is defined, the projected angle is uniquely determined. This is expressed in the following equation:

$$\Phi_{2D} = \text{render}(S) \qquad (1)$$

which states that projected angle (ranging from 0 to π) is uniquely determined by the scene that gave rise to the projection[1]. This mapping from a particular scene (i.e. a combination of Φ_{3D}, σ and ρ) onto the image is called the image formation function (Fig. 4a, b and c show the image formation functions when 3-D angle is 80°, 90°, and 100°, respectively).

From Fig. 4 it is clear that there is ambiguity about which scene gives rise to a particular projected angle. The projected angle can be a projection of a certain scene (Φ_{3D}, σ and ρ) but it can be a projection of another scene as well (with a different combination of 3-D angle, slant, and roll). For example, a projected angle of 103° can be the projection of a scene with an orthogonal angle and a 40° slant and a 60° roll (see inset of Fig. 4b) but it can also be the projection of a 3-D angle of 80° with a 49° slant and a 45° roll (see inset of Fig. 4a).

However, this ambiguity can be reduced by adding some depth information to the stimulus (i.e. by embedding the angle within a solid figure, see Section 2) and by the tendency to perceive the most regular interpreta-

---

[1] Note that the tilt of the angle relative to the plane of projection (i.e. the direction of the normal of the 3D angle in the image plane) does not appear in the equation because the projected angle does not change when altering the tilt of the angle.
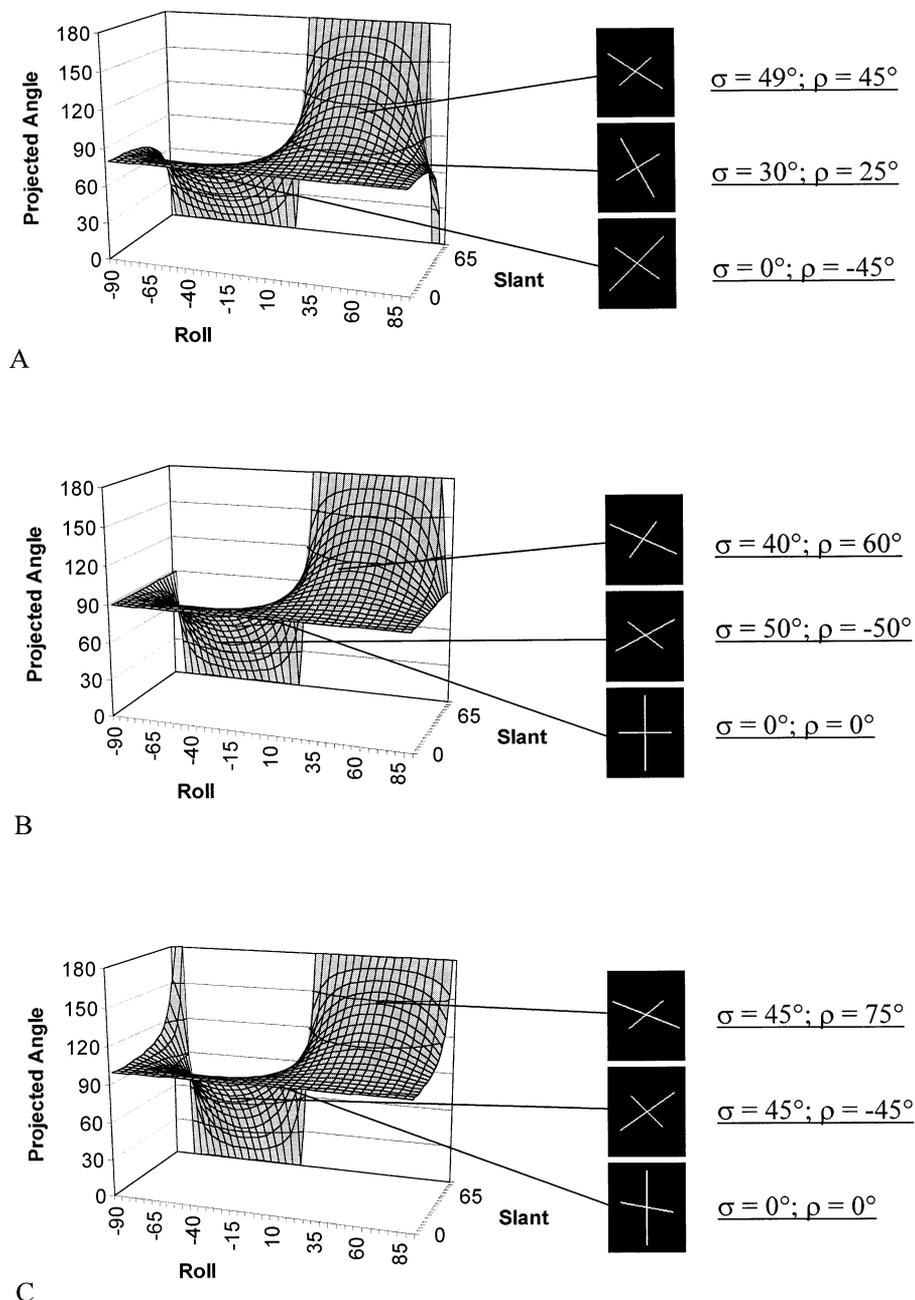
Fig. 4. The image formation functions with a fixed 3-D angle of 80° (a), 90° (b) and 100° (c). Note that each scene-vector is mapped onto one unique projected angle (Φ_{2D}) but that one projected angle is mapped onto several scene-interpretations.

tion (orthogonal interpretation). In principle, the additional depth information itself can be used to disambiguate the stimulus but estimation of the actual orientation is likely to be subject to some error (i.e. adding some depth information results in a distribution over the slant that is centred over the *perceived orientation* and not over the *objective orientation*).

Based on this formal task description, we modelled the stimulus information by means of a Bayesian observer. This approach seemed natural to us because the information in the images can be combined with a priori distributions over the relevant world properties

(e.g. Knill & Richards, 1996; Mamassian & Landy, 1998) resulting in a posterior probability distribution that can be viewed as the information that is provided by the images after taking into account the task faced by our observers and the a priori information of the relevant world properties.

Because judgements have to be made about the orthogonality of a 3-D angle, we want to compute this posterior probability (defined over all possible scenes) and use this information to calculate the probability of a certain projected angle to be orthogonal with a certain orientation (i.e. a scene):

$$P(\text{ortho}\,|\,\Phi_{2\mathrm{D}}) = \int_{\Omega} P(S\,|\,\Phi_{2\mathrm{D}})\,\mathrm{d}S \qquad (2)$$

The integration takes place over the domain Ω, which is the product of the domain for an orthogonal 3-D angle ($\pi/2 - \varepsilon_{3\mathrm{D}}$ to $\pi/2 + \varepsilon_{3\mathrm{D}}$), the domain for the slant of the angle (0 to $\pi/2$) and the domain for the roll of the angle ($-\pi$ to π).

In order to compute this value, we first have to define the likelihood for each interpretation (with S for scene). Then we model the prior probability for each scene (i.e. a 3-D density function defined over this space). These two measurements will then be combined using Bayes' rule (assuming that $\Phi_{3\mathrm{D}}$, $\sigma_{3\mathrm{D}}$ and $\rho_{3\mathrm{D}}$ are independent). The integrand of the above equation can then be rewritten:

$$P(S\,|\,\Phi_{2\mathrm{D}})\,\frac{P(\Phi_{2\mathrm{D}}|S)f(S)}{P(\Phi_{2\mathrm{D}})} \qquad (3)$$

where $P(\Phi_{2\mathrm{D}}|S)$ is the likelihood-function and $f(S)$ is the a priori density function defined over scene space.

The likelihood-function can be calculated by assuming Gaussian noise that is added while rendering a scene (Eq. (1)) and the a priori density function can be calculated by assuming a Gaussian distribution that is

a

b

Fig. 5. Model-output for the best-fit parameter-setting: p1 = 1.3, p2 = − 5.0, p3 = 4.0 and p4 = 0.34. (a) Effect of slant on the proportion of error for each type of object (averaged over the different roll-values). (b) Effect of roll on the proportion of error for each type of object (averaged over the different slant-values). Compare to Fig. 3 with similar psychophysical results.

centred on the orthogonal angle and the perceived orientation (see Appendix A for details).

Because we were unable to compute the integral of Eq. (2) analytically, we numerically integrated the posterior probability distribution. Further, we assumed that our subjects had perfect knowledge about the objective roll. The free parameters of the model were adjusted so as to provide a good fit between the observed probabilities and the probabilities calculated based on the posterior distribution (using the method of maximum likelihood).

Because we assumed that the judgements of our observers reflected these probabilities ($1 - P(\text{orthogonal})$ for targets and $P(\text{orthogonal})$ for distractors), we called this dependent variable the proportion of error of the model. These model error proportions are shown in Fig. 5.

Three results should be noted about the performance of the Bayesian model. First, it appears that the performance of the model is clearly viewpoint-dependent: The proportion of error of the model does not depend on the deviation of the distractor from orthogonality (because it was fixed over the different experimental trials) but depended on the orientation at which the distractor was viewed. Moreover, these errors were only observed when confronted with distractor objects. When confronted with a target object, the model shows a very low error rate, comparable with the low error rates obtained with human observers. Second, the error rate for the distractors was again linearly dependent on the slant of the angle relative to the plane of projection (see Fig. 5a). High values of slant resulted in high error rates. Finally, the error proportions for distractors were also dependent on the roll of the angle. When the 3-D angle was smaller than 90° (i.e. distractor 1), a positive roll resulted in lower error proportions than a negative roll, and vice versa when the 3-D angle was larger than 90° (i.e. distractor 2) (see Fig. 5b).

Moreover, the qualitative performance of the model (low error rate for targets, linear effect of slant, differential effect of the roll for the two distractors) was very robust against changes in parameter settings. This convinced us that the viewpoint-dependent performance of our human subjects could be expected, given the computational nature of the task at hand.

4. General discussion

We presented observers some projected angles and asked whether these angles were orthogonal in 3-D space or not. Because in the absence of additional information about the orientation of the angle relative to the plane of projection this task could not be solved, not even in principle, we embedded the angle of interest in a solid figure. The additional information at the Y-
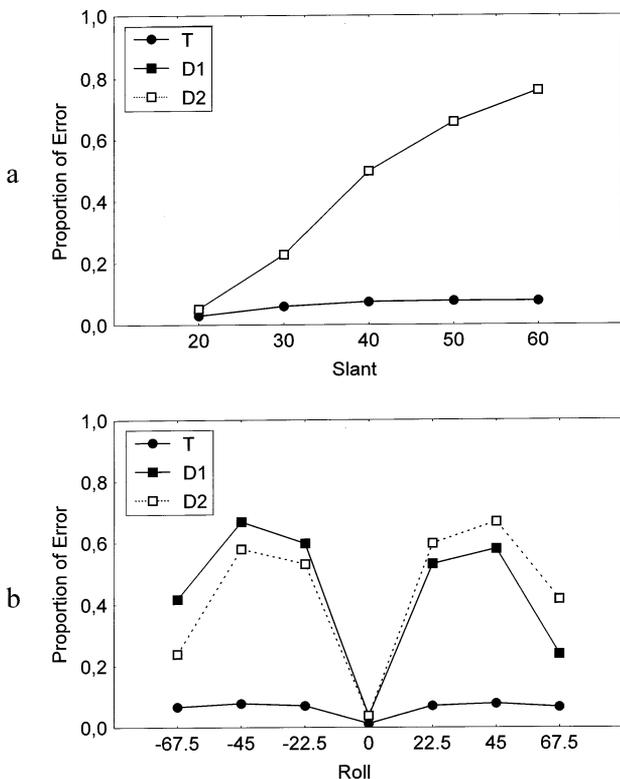
and arrow-junctions at the end-points of the legs of the crosses made it possible for observers to calculate the objective orientation in principle, so as to calculate which projected angle could be expected, if the 3-D angle were orthogonal.

However, we observed that some non-orthogonal angles were judged as being orthogonal and, moreover, that the probability of these non-veridical judgements depended on the orientation of these non-orthogonal angles relative to the observer. Whereas the fact that the errors were always observed for non-orthogonal angles could easily be explained by a bias toward the orthogonal interpretation, the viewpoint-dependency of these error proportions, at first sight, seemed to contradict a fixed set of constraints in order to reduce the set of possible interpretations.

In order to show that these viewpoint-dependent results do not exclude a fixed set of constraints that do not change when varying the viewpoint at which that particular scene is viewed, we calculated the posterior probabilities for each trial based on the image formation function of the task at hand, a simple model of the noise in the image formation process, and an a priori density function defined over the space of possible 3-D interpretations. The posterior probabilities of an implementation of this Bayesian model could be fitted to the observed probabilities using only four free parameters. From this it can be concluded that the viewpoint-dependency of our observers is a natural result, given the statistical information that is available in the images concerning the relevant world properties.

Viewpoint-dependent performance when judging 3-D objects from non-canonical viewpoints is often explained by an analogue process of aligning the incoming stimulus with a representation that is stored at some canonical viewpoint (e.g. Corballis, 1988; Jolicoeur, 1988; Ullman, 1989). However, the analysis of the statistical structure of the task faced by our observers shows that this viewpoint-dependency need not imply such an analogue mental transformation (see also King, Meyer, Tangney & Biederman, 1976; Wagemans, Van Gool & Lamote, 1996; Willems & Wagemans, 1999). A certain amount of uncertainty in the estimation of the orientation of the object towards the observer (i.e. the variance over and the underestimation of the slant) induced the viewpoint-dependent uncertainty in the judgements that had to be made.

It should be noted that, in the case of a distractor, there is a clear trade-off between the bias toward the orthogonal interpretation and the information present at the end-points of the legs of the crosses (two competing forces that 'pull' the interpretation in their direction). When confronted with a projected angle that is not the projection of an orthogonal angle ($\Phi_{3D} \neq 90°$), this projected angle can always be interpreted as an orthogonal angle (i.e. the likelihood function selects an orthog-

onal interpretation with a certain slant value as well as non-orthogonal interpretations with other slant values) but the probability of this to happen is dependent also on the a priori density, assigned to that interpretation. If the orthogonal interpretation (selected by the likelihood function) 'needs' a slant value that is far removed from the perceived slant, this orthogonal interpretation will have a low value, resulting in an oblique response (see Fig. 6 top row).

However, for certain other orientations, the interpretations selected by the likelihood function (including the orthogonal one), are situated closer to the mean of the a priori density function (i.e. it 'needs' a slant value closer to the perceived slant), resulting in a high error rate for that particular distractor at that particular orientation (see Fig. 6 bottom row).

From the computational analysis of the task, two conclusions have to be drawn with respect to the viewpoint-dependent performance of our observers. First, the viewpoint-dependency is a direct result of the fact that the information provided by the different images is changing over the different viewpoints (the posterior distribution shifts with respect to the orthogonal interpretations when changing viewpoint). Second, this change in information does not necessarily mean that the constraints imposed on the set of possible interpretations change with viewpoint. A priori distributions with fixed uncertainty were sufficient to induce a shift of the posterior distributions when changing the viewpoint at which the object was viewed.

Another assumption of this explanation needs some clarification. It is assumed that the uncertainty over the position of the crosses is fixed and does not depend on the orientation at which the cross is viewed (in terms of the model, the variance of the distributions over the orientation of the object is fixed and does not change with changing viewpoint). Some might argue that this uncertainty about perceived slant is not constant as a function of slant, especially in light of our own results presented in this paper (because the perceived slant is based on the projected angles at the end-points of the cross). However, this kind of reasoning is based on the assumption that only one end-point of the cross (a Y- or an arrow-junction) is used in order to calculate the perceived orientation. This is undoubtedly not the case. There are four end-points visible in the image (each with their own Y- and arrow-junctions) and each Y- or arrow-junction is viewed at a different orientation relative to the plane of projection. Because the perceived orientation is probably based on a global estimation (i.e. an estimation based on all the junctions visible in the image), we can assume that the slant variance is a constant function of the absolute slant at which the cross is viewed[2].

---

[2] In another recent study from our lab (Vanrie, Willems & Wagemans, 2000), we have obtained independent evidence that such a global estimation based on different angles considered in combination is possible.
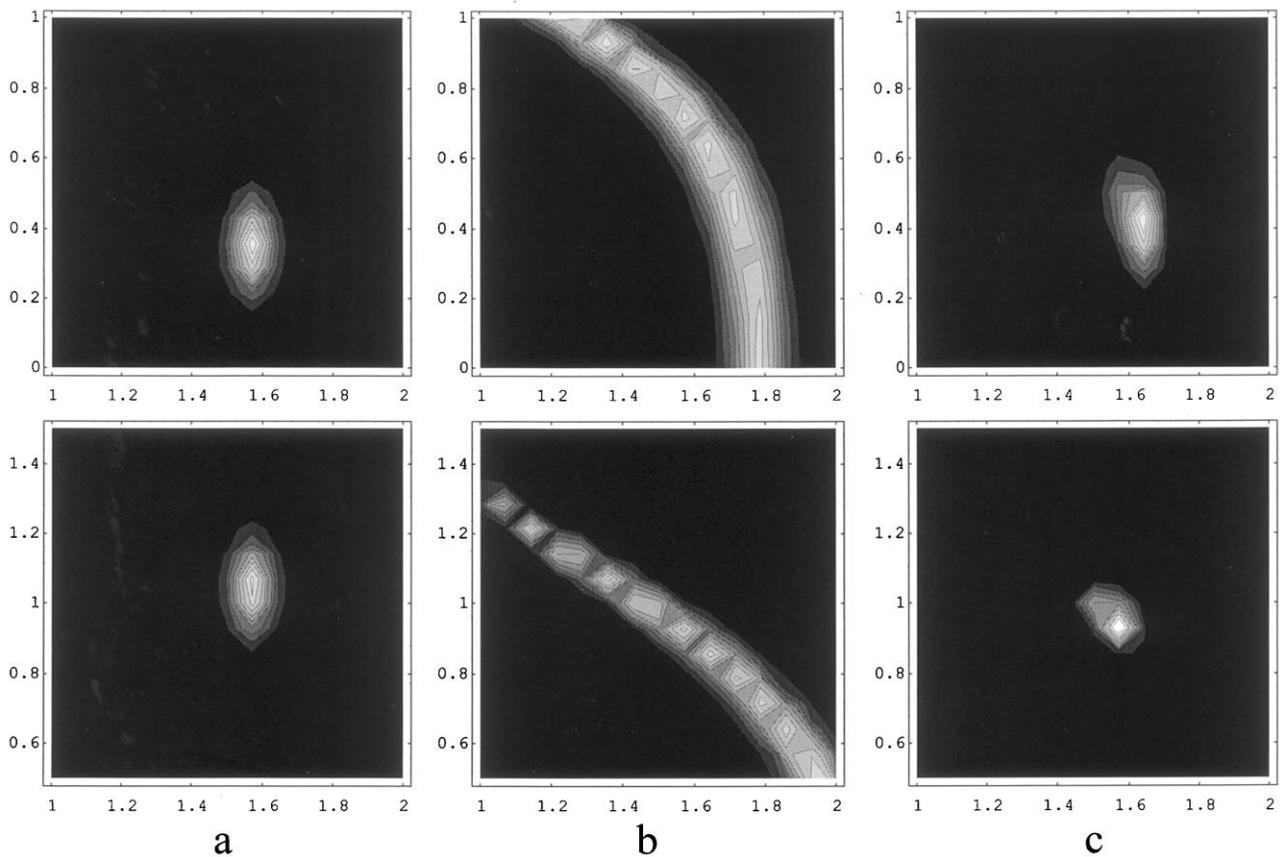
Fig. 6. Viewpoint-dependency in action. The top-row shows the model-output for a fixed set of parameter-values (p1 = 2.4, p2 = 0.0, p3 = 5.0, p4 = 3.0) for an 'easy' D2 trial (i.e. a trial with a low error proportion; $\sigma_o = 20°$, $\rho_o = 22.5°$, $\Phi_{2D} = 102°$) and the bottom row shows these terms for a difficult D2 trial (with a high error proportion; $\sigma_o = 60°$, $\rho_o = 45°$, $\Phi_{2D} = 118°$; values in the graph are expressed in radians). (a) The a priori density-function defined over scene-space (centred on the orthogonal angle with slant $\sigma_p = \sigma_o + p2$). (b) The likelihood-function for that trial (selecting those scenes that are compatible with Φ_{2D}). (c) Multiplication of the density-function and the likelihood-function (the numerator of Eq. (3)). The integration takes place over $\pi/2 - \varepsilon_{3D} < \Phi_{3D} < \pi/2 + \varepsilon_{3D}$ and $0 < \sigma < \pi/2$. Note how the likelihood-function shifts toward the mean of the a priori density-function for the difficult trial resulting in a higher density in the integrated domain (while the density-function is the same for the two trials).

Finally, because each component of the model (the uncertainty over the exact orientation of the object, the strength of the tendency to perceive the most regular interpretation, …) can be assigned a meaning (i.e. the role it plays in the process of inference), it could be interesting to see how these components change when changing the viewing conditions. Changing the viewing conditions can be understood in several ways. For example, it could be interesting to see how the components differ among different persons when doing this kind of task (i.e. the study of how the difference in performance of several observers is related to differences in underlying processes, assumed to play a role in the process of inference, each parameterised in our Bayesian model).

Another change of the viewing conditions could be the manipulation of additional depth information in the stimuli. For example, how does the uncertainty about the slant (i.e. parameter 2 and 3) vary when using for example stereo instead of the information contained in the Y- and arrow-junctions? How does this uncertainty change when these two kind of sources of depth information are combined in a single image? Does the noise of the likelihood function change when viewing the real scenes instead of scenes projected on a monitor? It is our belief that answers to these questions (questions that are already stimulating new research in our lab) can lead to significant insights in the process of interpreting 3-D objects from single images.

Acknowledgements

walle, Stefaan Tibau, Pedro Rosas, Tom Verguts and two anonymous reviewers for helpful comments on previous drafts.

Appendix A

In order to calculate the posterior distribution defined over scene-space (the vector composed of the 3-D angle Φ_{3D}, the slant σ, and the roll ρ), we first calculated the likelihood-function. In the absence of noise, this likelihood equals 1 for every scene that is compatible with the projected angle. It will be 0 otherwise (i.e. the likelihood is a function that selects those interpretations that can give rise to the projected angle, irrespective of the a priori probability of this interpretation). However, by assuming Gaussian noise that is added to the rendering[3] of a scene (mapping from scene to projected angle) the likelihood-function takes the following form:

$$P(\Phi_{2D}|S) = \frac{1}{\sqrt{2\pi\delta_n^2}} \exp\left(-\frac{(\Phi_{2D} - \text{render}(S))^2}{2\delta_n^2}\right)$$

So the likelihood decreases exponentially with the distance between projected angle and the angle obtained by rendering the scene and this drop is influenced by the noise δ_n of this likelihood-function.

Then this likelihood is weighted by the a priori density defined over the space of possible scenes (see Eq. (3)). Because we can assume that the different components of the scene are statistically independent we can calculate this 3-D density function by multiplying the three 1-D density functions. If we again assume Gaussian distributions over each component of scene space the 3-D angle prior distribution takes the following form:

$$P(\Phi_{3D}) = \frac{1}{\sqrt{2\pi\delta_{2D}^2}} \exp\left(-\frac{(\Phi_{3D} - \pi/2)^2}{2\delta_{3D}^2}\right)$$

This distribution is centered at the orthogonal angle with a strength of δ_{3D} (smaller values denote a stronger constraint).

The a priori distributions over the orientation of the projected angle then take the following form:

$$P(\sigma) = \frac{1}{\sqrt{2\pi\delta_\sigma^2}} \exp\left(-\frac{(\sigma - \sigma_p)^2}{2\delta_\sigma^2}\right)$$

$$P(\rho) = \frac{1}{\sqrt{2\pi\delta_\rho^2}} \exp\left(-\frac{(\rho - \rho_p)^2}{2\delta_\rho^2}\right)$$

By centering these distributions not on the objective slant and roll but on the perceived slant and roll, we left some room for a possible under- or overestimation of the exact slant and roll (see for example Cowie, 1998; Tibau, Willems, Van Den Bergh & Wagemans, 1999). This under- or overestimation of the exact slant and roll is parameterized by defining the perceived slant and roll in terms of the objective slant and roll:

$$\sigma_p = \sigma_o + \Delta\sigma$$

$$\rho_p = \rho_o + \Delta\rho$$

and the uncertainty of the orientation-estimation is then parameterized by the standard deviation of these Gaussian distributions (δ_σ and δ_ρ).

Using the above defined equations, we could calculate for each trial the likelihood and the a priori density for each scene. By integrating over the subspace of scene-space that could be considered orthogonal ($\pi/2 - \varepsilon_{3D} < \Phi_{3D} < \pi/2 + \varepsilon_{3D}$; $0 < \sigma < \pi/2$; and $-\pi < \rho < \pi$) and by using the fact that $P(\text{orthogonal}) + P(\text{non-orthogonal})$ should equal one (in this way it was not necessary to calculate the normalising constant of Eq. (3) explicitly), each trial could be assigned a probability of being interpreted as being orthogonal (ε_{3D} was fixed at 1°).

Further, we assumed that the judgements of the observers reflected this computed probability ('probability matching'; see also Mamassian & Landy, 1998). In the case of a target trial (i.e. a trial consisting of a depicted target object), the proportion of error should reflect $1 - P(\text{orthogonal})$, and in the case of a distractor object, the proportion of error should reflect $P(\text{orthogonal})$. In this way we could use the observed proportions of our observers for estimating the unknown parameters, using the method of maximum likelihood.

When implementing this theoretical model, we numerically integrated the space instead of analytical computing of the probabilities and we assumed that our subjects had complete knowledge about the objective roll of the angle (i.e. all density is concentrated on the objective roll)[4]. That way there were only four parameters left (δ_n, δ_{3D}, $\Delta\sigma$ and δ_σ) and these were adjusted so as to provide a good fit to the psychophysical data. The best fit was obtained with the following parameter values: $\delta_n = 0.95$, $\delta_{3D} = 1.36$, $\Delta\sigma = -4.8$ and $\delta_\sigma = 3.8$.

References

Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics*, 6, 391–396.

---

[3] Each scene was rendered by transforming (i.e. rolling and slanting) a pattern of three points with initial coordinates $(0, 1, 0)$ for the upper-point, $(0, 0, 0)$ for the mid-point, and $(\sin(x), \cos(x), 0)$ for the right-point (x is the deviation from orthogonality). After this transformation, we measured the angle between the three points as projected on the XY-plane.

---

[4] The estimation of this variance, based on a previous implementation of the model, resulted in a very low value, so we left out this component in order to keep the model tractable.

Autodesk, Inc. (1993). *Autodesk 3D studio (Release 3)*. Sausolito, CA: Author.

Cedrus, Corp. (1997). *Superlab Pro for Windows (Version 1.04)*. Phoenix, AZ: Author.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, *95*, 115–123.

Cowie, A. (1998). Measurement and modeling of perceived slant in surfaces represented by freely viewed line drawings. *Perception*, *27*, 505–540.

Griffiths, A. F., & Zaidi, Q. (2000). Perceptual assumptions and projective distortions in a 3-D shape illusion. *Perception*, *29*, 171–200.

Ittelson, W. H. (1952). *The ames demonstrations in perception*. Princeton, NJ: Princeton University Press.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, *42*, 461–478.

King, M., Meyer, G. E., Tangney, J., & Biederman, I. (1976). Shape constancy and a perceptual bias towards symmetry. *Perception and Psychophysics*, *27*, 129–136.

Knill, D. C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, *38*, 2635–2656.

Knill, D. C., & Kersten, D. (1991). Ideal perceptual observers for computation, psychophysics, and neural networks. In R. J. Watt, *Pattern recognition by man and machine* (pp. 83–97). Houndmills, UK: McMillan Press.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, MA: Cambridge University Press.

Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, *38*, 2817–2832.

Perkins, D. N., & Cooper, R. G. (1980). How the eye makes up what the light leaves out. In M. Hagen, *The perception of pictures*, *vol II. Durer's devices: beyond the projective model* (pp. 95–130). New York: Academic Press.

Tibau, S., Willems, B., Van Den Bergh, E., & Wagemans, J. (1999). The role of the centre of projection in the estimation of slant from texture of planar surfaces (submitted). Available as Internal Report No. 255, University of Leuven, Laboratory of Experimental Psychology.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition*, *32*, 193–254.

Vanrie, J., Willems, B., & Wagemans, J. (2000). Multiple routes to object matching from different viewpoints: mental rotation versus invariant features (submitted). Available as Internal Report No. 262, University of Leuven, Laboratory of Experimental Psychology.

Wagemans, J., Van Gool, L., & Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant. *Psychological Science*, *7*, 232–236.

Wallis, G. M., & Bülthoff, H. H. (1998). Using a 'virtual illusion' to put parallax in its place. 21st European Conference on Visual Perception, Oxford, UK (abstract). *Perception* (*Supplement*), *27*, 19a.

Willems, B., & Wagemans, J. (1999). Matching multi-component objects from different viewpoints: Normalization but not mental rotation (submitted). Available as Internal Report No. 236, University of Leuven, Laboratory of Experimental Psychology.