

Haptic perception disambiguates visual perception of 3D shape

Maarten W. A. Wijntjes · Robert Volcic ·
Sylvia C. Pont · Jan J. Koenderink ·
Astrid M. L. Kappers

Received: 19 November 2008 / Accepted: 12 January 2009 / Published online: 8 February 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract We studied the influence of haptics on visual perception of three-dimensional shape. Observers were shown pictures of an oblate spheroid in two different orientations. A gauge-figure task was used to measure their perception of the global shape. In the first two sessions only vision was used. The results showed that observers made large errors and interpreted the oblate spheroid as a sphere. They also misinterpreted the rotated oblate spheroid for a prolate spheroid. In two subsequent sessions observers were allowed to touch the stimulus while performing the task. The visual input remained unchanged: the observers were looking at the picture and could not see their hands. The results revealed that observers perceived a shape that was different from the vision-only sessions and closer to the veridical shape. Whereas, in general, vision is

subject to ambiguities that arise from interpreting the retinal projection, our study shows that haptic input helps to disambiguate and reinterpret the visual input more veridically.

Keywords Visual perception · Haptic perception · Cross-modal · 3D shape

Introduction

Different senses are frequently used for very similar purposes. This is particularly evident in the case of shape perception, which can be accomplished both in haptics and vision. When both senses are available, it is likely that they will play a complementary role. In the current study we wanted to investigate this sensory interaction for three-dimensional (3D) shape perception.

The visual image is ambiguous because the 3D environment is projected onto the two-dimensional (2D) retina. The brain needs to undo this projection, which is also known as the ‘inverse-optics’ problem: which shape, reflectance properties and light conditions could have caused the retinal image? Because of the underdetermination of this problem, image ambiguities arise. To resolve these ambiguities the brain makes use of a wide portfolio of computations that use so-called visual ‘cues’. Examples of cues are stereo, motion, texture, shading and contour-cues (Todd 2004). In many cases, multiple cues need to be available for a unique solution of the ‘inverse-optics’ problem. For instance, horizontal disparities between two retinal images are not sufficient (Mayhew and Longuet-Higgins 1982) but if motion is added the solution becomes unique (Richards 1985). Sufficient constraints do not guarantee unique percepts though (Todd 2004).

Electronic supplementary material The online version of this article (doi:10.1007/s00221-009-1713-9) contains supplementary material, which is available to authorized users.

M. W. A. Wijntjes (✉) · S. C. Pont
Faculty of Industrial Design Engineering,
Delft University of Technology, Delft, The Netherlands
e-mail: m.w.a.wijntjes@tudelft.nl

R. Volcic
Psychologisches Institut II, Westfälische Wilhelms-Universität,
Münster, Germany

J. J. Koenderink
Faculty of EEMCS,
Delft University of Technology, Delft, The Netherlands

A. M. L. Kappers
Physics of Man, Helmholtz Institute,
Utrecht University, Utrecht, The Netherlands

Interestingly, humans do not seem to be aware of these ambiguities. We are constantly experiencing a single solution and we are not consciously aware of ‘multiple visual worlds’ (Koenderink 2001).

Haptic perception of shape is rather different. While many haptic illusions exist (e.g. Hayward 2008), ambiguities due to projection are evidently not present. When a shape is probed by the fingers, the local shape and orientation are sensed by the mechanoreceptors, while the positions of the contact locations need to be encoded by muscle and joint receptors (kinaesthesia). The global shape can be perceived by sufficiently sampling these local inputs along the surface of the shape. Thus, the haptic channel has direct access to the shape, whereas the visual system needs to account for the projection transformations.

Two themes have dominated research on the interaction between vision and haptics. First, recognition of shapes has been studied (e.g. Newell et al. 2001; Norman et al. 2008). The main findings indicate that the two senses use partly similar but also partly different encoding principles. This makes the internal representation different and results in poorer recognition rates between senses than within senses. Second, it has been studied how sensory signals combine to form unitary percepts. When the inputs are in conflict, vision generally dominates (Rock and Victor 1964) but when the reliability of the visual signal is deteriorated, the haptic input receives increasingly more weight (Ernst and Banks 2002). These studies have been performed for length perception (Ernst and Banks 2002) and 2D shape perception (Helbig and Ernst 2007). The projection problems that the visual system needs to solve are evidently not present in these low-dimensional stimuli. Therefore, we wanted to investigate how haptics can influence the visual perception of a 3D shape.

One of the best available methods to probe the visual depth inference is a gauge-figure task described by Koenderink et al. (1992): an observer adjusts an ellipse so that it appears as a circle lying on the surface of the stimulus (see Fig. 1). This essentially provides the experimenter with subjective local attitudes along various surface positions. These data can be converted into depth maps, which reveal how observers infer the third dimension (depth) from a 2D image. We designed an experiment in which the visual stimulus could simultaneously be viewed and touched. We measured how the subjective relief of a 3D shape depends on the availability of haptic input. One of the earlier studies in visual perception of 3D shape made use of ellipsoidal stimuli (Mingolla and Todd 1986). They reported that observers were biased to interpret the shape as if the major axes were aligned with the picture plane. Therefore, we used

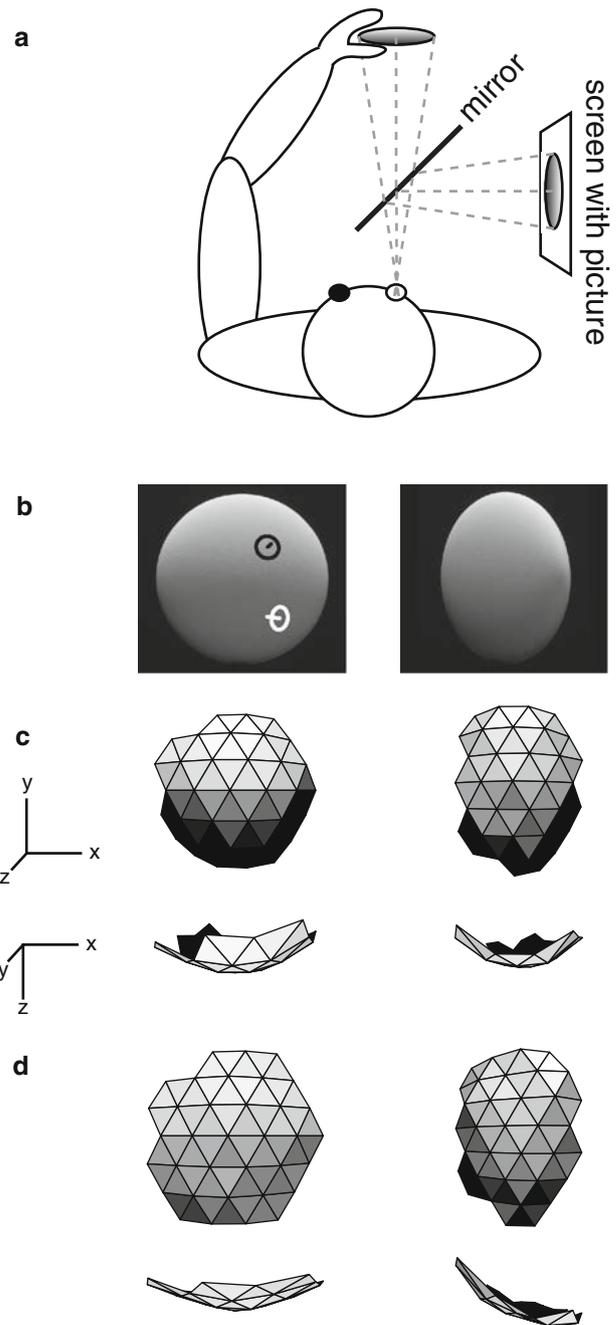


Fig. 1 **a** Top view of experimental setup. The left hand explored the stimulus. The mirror allowed to juxtapose the visual stimulus on the haptic location. Viewing was monocular, the left eye was blocked. **b** The visual stimulus in the frontal (*left*) and oblique (*right*) position. On the left, two gauge-figures are depicted: the *black* indicates a ‘proper’ setting whereas the *white* is clearly wrong. **c** Frontal and top views of depth maps of observer MB for the vision-only condition. **d** Similar to **c**, but for the vision + haptics condition

an ellipsoid and presented it in two different orientations. Both a general depth stretch and an obliquely oriented major axis could potentially lead to visual errors, which in turn could benefit from haptic input.

Methods

Participants

Four volunteers (2 males and 2 females), who had normal or corrected-to-normal visual acuity, participated in this experiment. They were unfamiliar with the purpose of the research.

Materials and apparatus

The stimulus was an ellipsoid with main axes of $10 \times 10 \times 3$ cm. For production a 3D printing technique (high quality stereolithography epoxy resins, precision of 0.1 mm) was used. Two similar versions were made and both were smoothed with sand paper. The visual stimulus was spray-painted matte white and the haptic stimulus was left unchanged.

The visual stimulus was photographed with a Canon EOS 400D. A white balance gauge was used to photometrically calibrate the images. To keep the visual experience as realistic as possible, the distance between the lens and the stimulus was similar to the distance between the observer and the screen, while using a neutral zoom of 50 mm. Thus, there were in principle proper perspective cues available to the observer (which could be relevant in the oblique condition). Because we did not want the observers to use the photographic ‘depth of field’ as a cue, we used the smallest aperture size (i.e. largest f-number) available: F20. This results in a picture that is equally focused throughout the depth range of the stimulus. The pictures were taken in a studio with black painted walls. The stimulus was placed directly below the light source, which was a row of fluorescent lights suspended on the ceiling.

During the experiment, the haptic stimulus was fixed on a standard and could be touched freely. As can be seen in Fig. 1, a mirror construction was used so that the position of the haptic stimulus coincided with the position of the visual stimulus. A chin rest was used and viewing was monocular; vision through the left eye was blocked. The observer could only see the image on the screen and not the haptic stimulus. Viewing distance was 40 cm, which makes the visual stimulus subtending an angle of 14° .

Procedure

The experiment consisted of four sessions. In the first two sessions the observers were presented with only the visual stimulus. In the last two sessions the observers performed essentially the same task, but also touched the haptic stimulus (see Fig. 1a). Two orientations of the stimulus were used: frontal (the short main axis of the ellipsoid was

normal to the picture plane) and oblique (45° rotated) as can be seen in Fig. 1b. The orientation condition was counterbalanced between the observers. Each session took approximately 15 min. The four sessions were completed subsequently with short breaks in between.

In each session, the observers had to perform a gauge-figure task: adjust an ellipse so that it appeared as a circle lying on the apparent surface of the stimulus. The gauge-figure is the projection of a circular disc with an orthogonally placed rod in the centre. Two examples of the gauge-figure can be seen on the left side of Fig. 1b: the black example illustrates a ‘proper’ setting and the white example clearly shows an ‘erroneous’ setting. Observers used the mouse with their right hand and adjusted the attitude of the gauge-figure. A triangular grid was used over which the gauge-figure was shown in random order. In the frontal condition the total number of trials (n) was 50 and in the oblique condition 49. The output consisted of slant and tilt pairs (σ_i, τ_i) that define the subjective orientation of the surface on the triangular grid (x_i, y_i) . These data can thus be used to reconstruct a depth profile (x_i, y_i, z_i) (see Koenderink et al. 1992 for details of this procedure). This process can be compared with integration: if only the derivative (orientation) is known of a certain function, the original function (height profile) can be reconstructed by integration. Similar to mathematical integration, there is an unknown constant of integration, which in our case is the absolute depth. For the purpose of our study, absolute depth is irrelevant. The total number of reconstructed depth coordinates z_i is 35 for both conditions (less than n because of triangulation). In Fig. 1c, two examples (each with frontal and top view) of a depth map are shown.

Data analysis

Comparison within subjects

We wanted to analyse the difference between the vision-only and vision + haptics condition. To analyse the differences between the two conditions we linearly regressed the depth values of both conditions with each other. Let z_i^v and z_i^{v+h} be the depth profiles for the vision-only and vision + haptics condition, respectively, then the following regression was performed:

$$z_i^{v+h} = \delta + \zeta z_i^v + \xi x_i \quad (1)$$

The regression coefficient δ is meaningless since it accounts for the arbitrary depth offset that results from the gauge-figure procedure. The coefficient ζ indicates the depth gain between the two conditions: if the vision-only condition results in a higher relief surface then $\zeta < 1$, and vice versa. In the example data on the left side of Fig. 1c, d, it can be seen that in this case $\zeta < 1$. We tested whether ζ

would statistically differ from 1 instead of from 0 since the former is more informative about whether the reliefs differ between the conditions. The last coefficient (ζ) accounts for a possible depth shear, or a so-called ‘additive plane’. In theory, this shear could have any orientation, but we have restricted the model to only one direction because of the symmetry of our stimulus. The meaning of this coefficient can best be interpreted when looking at Fig. 1c and d. In these depth maps the x -axis goes from left to right. For the oblique condition (right side) the two maps seem to be rotated with respect to each other. This is an affine shear transformation (which could also be called ‘affine rotation’) and can be modelled by the regression along the x -direction. We predict that ζ will be 0 for frontal condition and 1 (relating to the 45° orientation) for the oblique condition.

Comparison with veridical

We also wanted to compare the raw settings of the gauge-figure with respect to the veridical shape. The shape is defined by the solution of $f(x, y, z) = 0$ for

$$f(x, y, z) = \left(\frac{x}{r_x}\right)^2 + \left(\frac{y}{r_y}\right)^2 + \left(\frac{z}{r_z}\right)^2 - 1 \quad (2)$$

where r_x, r_y, r_z are the three main ellipsoid radii. Furthermore, a rotation θ around the y -axis is part of the model. We fixed r_y at 1 because it is defined by the height of the stimulus in the picture. For each position in the triangulation we calculated the normal vector of the shape (Eq. 2) by $\mathbf{n} = \nabla f(x, y, z)$. The subjective normal vectors are defined by the slant σ and tilt τ : $\mathbf{n}_s = (\cos \tau \sin \sigma, \sin \tau \sin \sigma, \cos \sigma)$. The fitting procedure aimed to minimise the inner product between the model and subjective normal vectors $\langle \mathbf{n}, \mathbf{n}_s \rangle$. We used a nonlinear regress procedure to minimise this function with respect to the parameters (r_x, r_z, θ) . Veridical values would be (1, 0.3, 0) and (1, 0.3, $\pi/4$) for the frontal and oblique conditions, respectively.

Results

Comparison between and within subjects

To assess the similarity between the four observers we correlated the depth values of the reliefs between the observers within each condition. The average correlation was $r = 0.92$ with a lowest value of $r = 0.78$ and a highest value of $r = 0.98$. Because of this high consistency it makes sense to first look at the ‘raw’ data of observer MB in Fig. 1c, d. A frontal and top view are available for each condition. On the left side, the ‘frontal’ data are shown. It can be clearly seen that the stimulus was perceived to be

much more curved (higher relief surface) when there was no haptic information available. On the other hand, when observers could touch the stimulus (Fig. 1d), they perceived the stimulus as flatter (and closer to veridicality). In the ‘oblique’ vision-only condition, the observers seemed to perceive a kind of egg shape (prolate spheroid), symmetrical around the y -axis. When the observers could explore the stimulus haptically, their percept seemed to change towards the veridical disc shape (oblate spheroid). Depth maps of all observers can be found in Supplementary material.

The regression coefficients defined in Eq. (1) are presented in Table 1. The depth stretch coefficients (ζ) are all significantly below 1, reflecting that observers perceived the stimulus flatter when the stimulus could be touched. Furthermore, the affine rotation parameter (ξ) is nearly 0 in the frontal condition, whereas it is clearly negative in the oblique condition. This shows that observers perceived a differently oriented stimulus when touch was available. Note that the coefficient ξ is clearly non-zero. However, ξ was also below 1, the value that relates to a 45° orientation difference.

Comparison with veridical

To represent the 3D parameter data (r_x, r_z, θ) comprehensively, we used a polar coordinate transformation where the radius (distance from origin) is represented by (r_z/r_x) and θ is the polar angle. Note that the reciprocal of the axes ratio corresponds to a phase shift of 90° : a stimulus of $(r_x, r_y, r_z) = (10, 10, 3)$ is similar to $(r_x, r_y, r_z) = (3, 10, 10)$ rotated 90° around the y -axis. Therefore, we took the reciprocal value of (r_z/r_x) if it exceeded 1 and applied the phase shift $\theta \rightarrow \theta + \pi/2$. Furthermore, since the shape is invariant under 180° rotations we projected all data to the $[0^\circ, 180^\circ)$ interval.

Table 1 Results of the regression within observers and between the depth profiles from the different conditions (see Eq. 1)

Observer	Condition	ζ ($p_{\zeta=1}$)	ξ ($p_{\xi=0}$)
MB	Frontal	0.45 (<0.0001)	-0.02 (0.2876)
AD	Frontal	0.61 (<0.0001)	-0.02 (0.2108)
BH	Frontal	0.31 (<0.0001)	-0.02 (0.0016)
AB	Frontal	0.26 (<0.0001)	0.00 (0.5698)
MB	Oblique	0.73 (<0.0001)	0.37 (<0.0001)
AD	Oblique	0.39 (<0.0001)	0.44 (<0.0001)
BH	Oblique	0.35 (<0.0001)	0.25 (<0.0001)
AB	Oblique	0.55 (<0.0001)	0.71 (<0.0001)

Parameter ζ denotes the depth scaling along the viewing direction; a value of 1 would indicate no scaling. Parameter ξ indicates whether there is an orientation difference between the conditions

The result is shown in Fig. 2. The thick semicircles at $(r_z/r_x) = 0.3$ and $(r_z/r_x) = 1$ represent ellipsoids with the veridical axes ratio and the axes ratio of a sphere, respectively. Note that points that are near the $(r_z/r_x) = 1$ locus are rotationally symmetrical and therefore the value of θ becomes meaningless (a sphere is rotation invariant). The black data represent the fitted shapes from the vision-only sessions and the grey data from the vision + haptics sessions. The black crosses represent the veridical stimulus. As can be seen, all vision-only data are positioned further towards the spherical shape $(r_z/r_x) = 1$ than the vision + haptics data. Furthermore, the orientation is more veridical in the vision + haptics condition. Note that the rotations (θ) of the vision + haptics data from the oblique condition are all $<45^\circ$. This means that although the observers perceived that the stimulus was obliquely oriented, they underestimated the amount of rotation. This result relates to the findings in the previous section that the ξ coefficients were all below 1.

Discussion

The results show that haptic exploration can influence and improve visual perception of 3D shape. All observers perceived a shape that was less spherical and less oriented towards the picture plane when haptic information was added (Table 1). It was also shown that this percept was

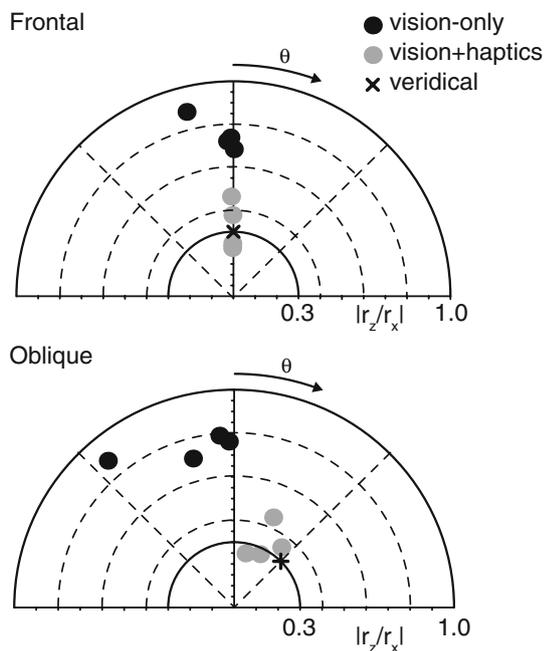


Fig. 2 Polar plot representations of the results. The veridical shape is indicated by the cross. The angle θ represents the orientation and the radius represents (r_z/r_x)

closer to the veridical shape (Fig. 2). Furthermore, it can be seen in this figure that, although the vision-only condition is ambiguous, there still seems to be systematicity: all data points are around an axis ratio of about 0.8 and in roughly similar directions. Also the correlation between subjects was rather high. This indicates that observers resolve the ambiguity in an erroneous but surprisingly similar way. It may be tentative, but the spread in the shape parameters seems to be similar in the vision-only and the vision + haptics conditions (taking into account that it is a polar plot). This could mean that observers do adjust their percept towards the veridical shape, but preserve some idiosyncratic differences. What can also be observed from the vision + haptics data is that the orientation seems to be biased (all less than 45°) but that the axes ratio is spread around 0.3. In the context of sensory integration (Ernst and Bühlhoff 2004) it thus seems that the final orientation is based on an average of the two senses, whereas the depth stretch is completely dominated by the haptic input. In other words, curvature (second order shape information) is completely captured by haptics, whereas overall orientation (first order shape information) is integrated between the two senses. This latter finding has also been shown in the work of Ernst et al. (2000).

The orders of the vision-only and vision + haptics conditions were fixed, so one could argue that the effect is due to learning. The reason we did not perform the same experiment in reversed order is that this would elicit novel questions that are beyond the scope of this research. If observers participated first in a vision + haptics condition and subsequently in a vision-only condition, the latter session could show traces related to the previously touched stimulus. The results should then be interpreted in this memory-related context, which is certainly an interesting topic, but not in line with the current research question. Furthermore, we have strong indications that learning effects have played a negligible role in our experiment. First, the orientation condition was counterbalanced and there did not appear to be clear differences with respect to the order (see also Supplementary material). Second, it has been found in earlier studies that there are high correlations within subjects performing the same task multiple times (Koenderink et al. 2001). This implies that observers do not change their percept from one session to the other and that the change reported in this study is due to haptic input.

The visual stimulus was a photograph of the original shape. One could argue that using the actual object as visual stimulus would be more ecological valid. The problem with using real visual objects in these kinds of experiments is that the gauge-figure needs to be projected onto the stimulus. This can be accomplished by using a laser system that renders the gauge-figure onto the stimulus (Koenderink et al. 1995). Perception of real 3D shapes is a

relatively undeveloped but relevant topic for future research. However, this was beyond the scope of the current study.

The present study shows that the haptic sense complements visual perception of 3D shape. Touch seems to recalibrate the visual system so that it is better able to infer depth from the retinal projection. This is reminiscent of the “touch educates vision” idea of Berkeley (1963/1709). Although this issue still elicits debate there is no reason to doubt that the visual image is ambiguous and that touch improves visual perception. We can thus rephrase Berkeley’s thought to “touch disambiguates vision”.

Acknowledgments This research was supported by grants from the Netherlands Organisation for Scientific Research (NWO) and a grant from the EU (FP7-ICT-217077-Eyeshots).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Berkeley G (1963/1709) A new theory of vision and other writings. Dutton, New York
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870): 429–433
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8(4):162–169
- Ernst MO, Banks MS, Bühlhoff HH (2000) Touch can change visual slant perception. *Nat Neurosci* 3(1):69–73
- Hayward V (2008) A brief taxonomy of tactile illusions and demonstrations that can be done in a hardware store. *Brain Res Bull* 75:742–752
- Helbig HB, Ernst MO (2007) Optimal integration of shape information from vision and touch. *Exp Brain Res* 179:595–606
- Koenderink JJ (2001) Multiple visual worlds. *Perception* 30(1):1–7
- Koenderink JJ, van Doorn AJ, Kappers AML (1992) Surface perception in pictures. *Percept Psychophys* 52:487–496
- Koenderink JJ, van Doorn AJ, Kappers AM (1995) Depth relief. *Perception* 24(1):115–126
- Koenderink JJ, van Doorn AJ, Kappers AML, Todd JT (2001) Ambiguity and the ‘mental eye’ in pictorial relief. *Perception* 30(4):431–448
- Mayhew JEW, Longuet-Higgins HC (1982) A computational model of binocular depth perception. *Nature* 297(5865):376–378
- Mingolla E, Todd JT (1986) Perception and solid shape from shading. *Biol Cybern* 53(3):137–151
- Newell FN, Ernst MO, Tjan BS, Bulthoff HH (2001) Viewpoint dependence in visual and haptic object recognition. *Psychol Sci* 12(1):37–42
- Norman JF, Clayton AM, Norman HF, Crabtree CE (2008) Learning to perceive differences in solid shape through vision and touch. *Perception* 37:185–196
- Richards W (1985) Structure from stereo and motion. *J Optical Soc Am A Optics Image Sci Vis* 2(2):343–349
- Rock I, Victor J (1964) Vision and touch: an experimentally created conflict between the two senses. *Science* 143(3606):594–596
- Todd JT (2004) The visual perception of 3D shape. *Trends Cogn Sci* 8(3):115–121