Reaching into Pictorial Spaces

Robert Volcic[a], Dhanraj Vishwanath[b] and Fulvio Domini[a,c]

[a] Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy; [b] School of Psychology & Neuroscience, University of St. Andrews, St. Andrews, UK; [c] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA

ABSTRACT

While binocular viewing of 2D pictures generates an impression of 3D objects and space, viewing a picture monocularly through an aperture produces a more compelling impression of depth and the feeling that the objects are "out there", almost touchable. Here, we asked observers to actually reach into pictorial space under both binocular- and monocular-aperture viewing. Images of natural scenes were presented at different physical distances via a mirror-system and their retinal size was kept constant. Targets that observers had to reach for in physical space were marked on the image plane, but at different pictorial depths. We measured the 3D position of the index finger at the end of each reach-to-point movement.

Observers found the task intuitive. Reaching responses varied as a function of both pictorial depth and physical distance. Under binocular viewing, responses were mainly modulated by the different physical distances. Instead, under monocular viewing, responses were modulated by the different pictorial depths. Importantly, individual variations over time were minor, that is, observers conformed to a consistent pictorial space. Monocular viewing of 2D pictures thus produces a compelling experience of an immersive space and tangible solid objects that can be easily explored through motor actions.

Keywords: pictorial space, vision, 3D, depth perception, monocular, binocular, motor action, reaching

1. INTRODUCTION

Pictorial images depicting a 3D scene generate an unambiguous and evidently undistorted perception of 3D space and objects.[1–5] However, the qualitative experience of pictorial space under normal viewing is different from that obtained when the original real scene is viewed with both eyes.[2,6–11] Viewing pictorial images normally with one or two eyes does not produce the same qualitative vividness—the sense of immersive space and tangible solid objects—that is obtained under binocular viewing of the real scene. This difference persists even when the image produced in the eye is perspectivally the same in the two cases,[2,12] although haptic exploration can influence and improve visual perception of 3D shape.[13,14] This deficiency of pictorial depth perception is usually attributed to the lack of parallax information (binocular disparities or motion parallax) when viewing a flat picture.[6,15–17] Parallax information is thought to not only produce a qualitatively vivid perception of three-dimensionality but to also provide high precision information that can be used to control reaching and grasping movements.[18,19]

Recently it has been empirically demonstrated that viewing a picture monocularly through a reduction aperture can generate the same vivid qualitative attributes of three-dimensionality associated with binocular stereopsis or parallax,[11] confirming previous introspective accounts (e.g., Brunelleschi and Da Vinci (cited in Ref. 12) and Refs. 2, 7, 8). Here we asked if this mode of viewing pictures contributes to an enhancement in visually guided motor control similar to that observed in the presence of parallax information. To do this we compared the capacity for subjects to point in depth within a virtual pictorial space under both binocular and monocular viewing. In both cases the pictorial image was viewed through apertures that occluded its rectangular boundaries. We selected natural images with rich pictorial depth structure depicting objects of a visual scale and distance that were plausibly consistent with being located in reach space (Figure 1, top panels). Images were displayed on a CRT computer monitor using a mirror setup that allowed free movement of the hand without visual feedback of hand position (Figure 1, bottom panels). We varied both the location in pictorial space that

Corresponding author: Robert Volcic: E-mail: robert.volcic@iit.it, Telephone: +39 0464430161

Human Vision and Electronic Imaging XIX, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Huib de Ridder, Proc. of SPIE-IS&T Electronic Imaging, Proc. of SPIE Vol. 9014, 901413
© 2014 SPIE-IS&T · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2045458

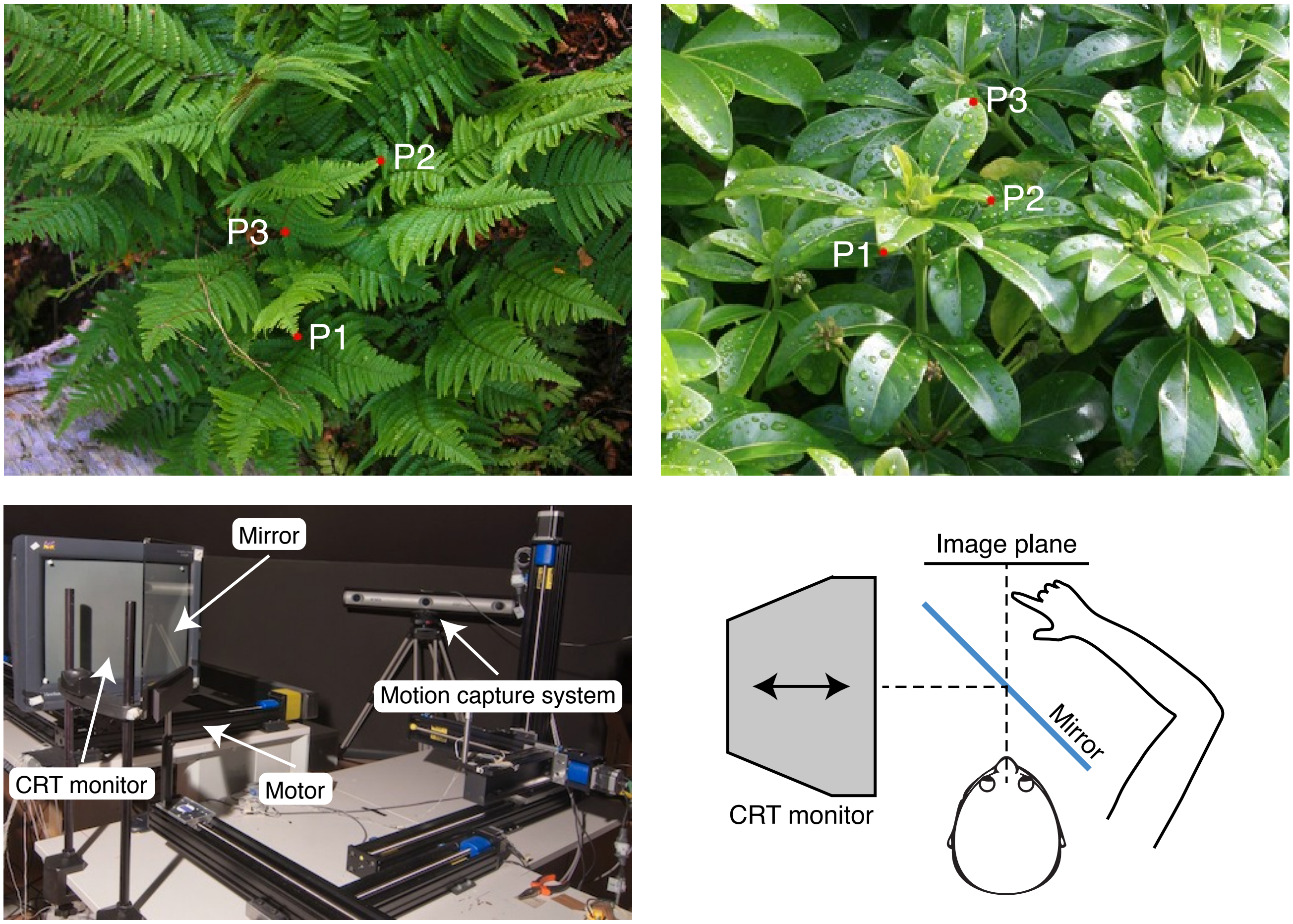

Figure 1. Top, photographic images of natural scenes used as stimuli (left - ferns, right - leaves). In each image three locations were selected which correspond to different pictorial depths (near - P1, middle - P2, far - P3). These locations served as targets in the reach-to-point task. On each trial, subjects saw one of the two images with only one of the red targets. Bottom left, a picture of the setup showing the arrangement of the motion capture system, the mirror in front of the monitor and the motor for moving the monitor. Bottom right, a schematic top view of the setup. Subjects pointed in depth under either monocular or binocular viewing. The distance of the image plane on which the stimulus was presented was adjusted by moving the monitor.

was to be pointed to (Figure 1, top panels, red targets), as well as the distance of the picture surface, i.e., the image plane (distance of the CRT monitor). Finger position during pointing was measured using a motion capture system.

We found that pointing in depth was better modulated by virtual pictorial depth under monocular aperture viewing compared to binocular aperture viewing. Under binocular viewing, differences in pointing responses were attributable mostly to changes in physical monitor distance. This suggests that monocular aperture viewing affects visually guided motor actions in a way similar to that observed under binocular stereopsis.

2. METHODS

2.1 Participants

Six undergraduate students (3 females) participated in this study. All had normal or corrected-to-normal vision. All of the subjects were naïve to the purpose of the experiments and were paid for their effort. Experiments were undertaken with the understanding and written consent of each subject, with the approval of the Comitato Etico per la Sperimentazione con l'Essere Vivente of the University of Trento, and in compliance with national legislation and the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki).

2.2 Apparatus

Subjects were seated in a dark room in front of a high-quality, front-silvered 400 × 300 mm mirror (Figure 1, bottom panels). The mirror was slanted at 45° relative to the subjects' sagittal body mid-line and reflected the image displayed on a ViewSonic 9613, 19" CRT monitor placed directly to the left of the mirror. For consistent vergence and accommodative information, the position of the monitor, attached to a linear positioning stage (Velmex Inc., Bloomfield, NY, USA), was adjusted on a trial-by-trial basis to equal the distance from the subjects' eyes to the image plane. Subjects' head were stabilized using a chin rest. A custom C++ program was used for stimulus presentation and response recording.

Movements of the index finger were acquired on-line at 100 Hz with sub-millimeter resolution by using an Optotrak Certus motion capture system with two position sensors (Northern Digital Inc., Waterloo, Ontario, Canada). The position of the tip of the finger was calculated during the system calibration phase with respect to three infrared-emitting diodes attached on the distal phalanx. Additional details about the setup are available in the companion article.[20]

Stimuli were color photographs of natural scenes (see Figure 1, top panels). In each image we selected three locations which correspond to different pictorial depths (near - P1, middle - P2, far - P3). These locations served as targets in the reach-to-point task. Targets were red discs and were superimposed on the image, one at a time. The images were presented along the line of sight at two distances (420 and 520 mm from the subject) and they were scaled to be retinally identical. When displayed at the closest distance, the width and the height were 250 and 187.5 mm, respectively. When displayed at the farthest distance, the width and the height were 309.5 and 232.1 mm, respectively. Subjects wore a pair of goggles with two circular apertures (radius: 10 mm). Under binocular viewing, both apertures were kept open, whereas, under monocular viewing, the aperture in front of the left eye was occluded. An additional aperture (radius: 62.5 mm) was positioned 400 mm from the subjects to block the view of the monitor frame.

The setup allowed subjects to comfortably reach behind the mirror to perform reaching movements with their right hand (Figure 1, bottom panels). The hand starting position (a 260 mm high pole) was shifted relative to the body of the observer by about 250 mm from the coronal plane and 150 mm from the sagittal plane. No visual feedback about the hand position was available during reaching. A keyboard was positioned in front of the subjects to register their responses.

Reach-to-point movements consisted of 120 trials per subject resulting from 2 viewing modes (monocular and binocular) × 2 monitor distances (physical distance) × 3 targets (pictorial depth) × 2 images × 5 repetitions. These trials were in pseudo-randomized order, but were blocked according to the viewing mode. The experiment took on average approximately 30 minutes.

2.3 Procedure

Subjects were tested in a dark room with their head positioned on a chin rest to maintain the same head position during all blocks. Before starting the experiment subjects were presented with a set of practice trials to get accustomed to the task. Subjects started each trial of the experiment with their index fingertip resting on the top of a pole. Subjects were instructed to align their invisible index finger of their right hand with the position in space in which the target was located. Once their finger was in position, they had to press with their left hand the space bar on the keyboard to confirm their response. After the response the monitor turned black and the subjects returned with their hand to the starting position. Then, the monitor moved to the new position ready for the start of the next trial. The experiment was completed in one testing session that consisted of two blocks of trials interrupted by a short pause.

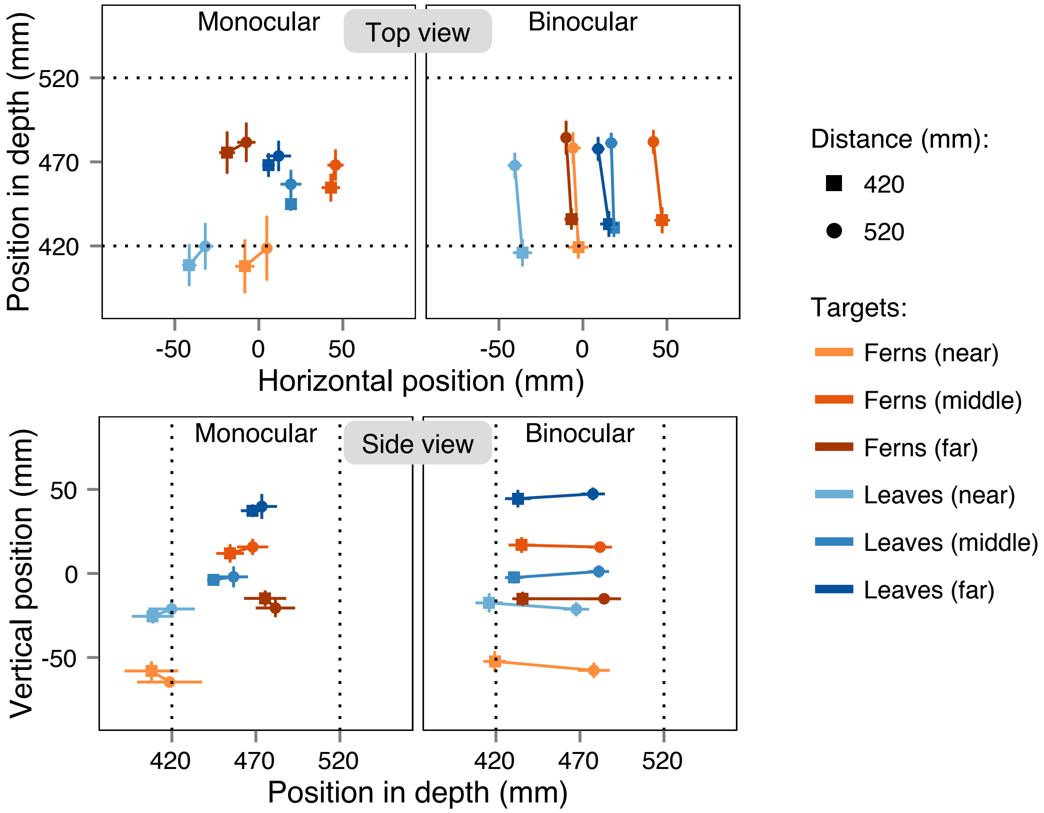

Figure 2. Top and side views of the average index position (± s.e.) in monocular and binocular conditions for all the targets. The responses to the same target, but for different physical distances are connected by a continuous line within each panel. Horizontal (top panels) and vertical (bottom panels) dotted lines indicate the physical distances at which the images were presented.

2.4 Data Analysis

The raw positional data were smoothed and differentiated with a 2nd order Savitzky-Golay filter with a window size of 41 points. These filtered data were then used to compute the velocity in 3D space of the index finger tip. Trials in which the end of the reaching movement could not be identified correctly (e.g., the hand kept drifting) were excluded from further analyses (2.6%). All analyses were performed in in R statistical programming language[21] and the figures were generated with the *ggplot2* package.[22]

3. RESULTS

The average index position for all conditions is shown in Figure 2. A clear distinction in response patterns between viewing modes is already evident. First, the pictorial depth domain under monocular viewing was more expanded than under binocular viewing. Second, the responses under binocular viewing were more strongly affected by the physical distance of the image than under monocular viewing. However, physical distance modulated the binocular responses only partially: responses were close to the veridical distance when the image was at the closer distance, but they were slightly more than halfway when the image was presented at the farther distance.

To analyze the data, we ran two separate repeated measures ANOVA, one for each stimulus image. We analyzed how the index finger position in depth varied as a function of viewing mode (monocular and binocular), physical distance (420 and 520 mm) and pictorial depth (near, middle and far).

Both analyses yielded very similar results. We found a significant main effect of pictorial depth (Ferns: $F(2,10) = 6.963$, $p = 0.013$; Leaves: $F(2,10) = 14.57$, $p = 0.001$) and physical distance (Ferns: $F(1,5) = 40.7$, $p = 0.001$; Leaves: $F(1,5) = 34.77$, $p = 0.002$). The main effect of condition was not significant (Ferns:

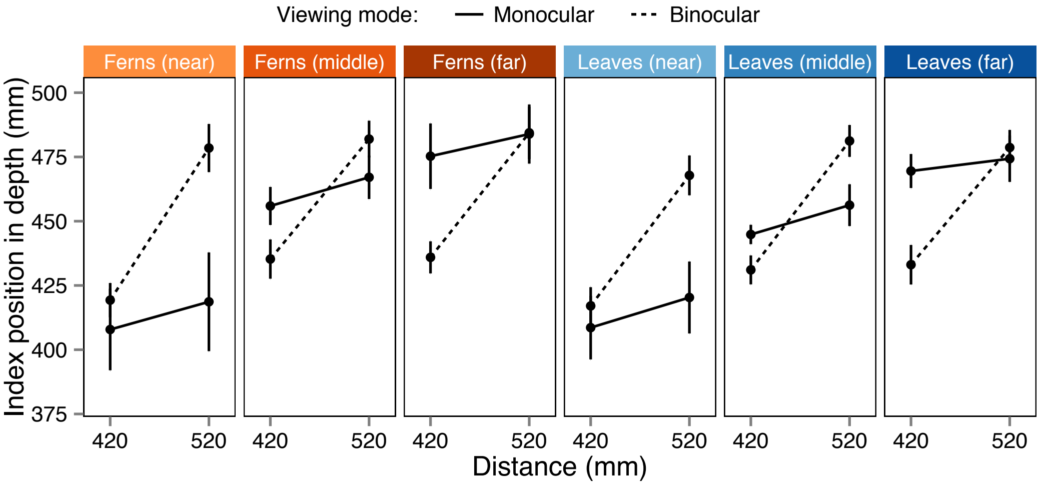

Figure 3. Average index position in depth (± s.e.) for all targets. Each panel summarizes the responses to the same target as a function of physical distance and viewing mode. Color coding is as in Figure 2.

$F(1,5) = 0.215$, $p = 0.66$; Leaves: $F(1,5) = 0.461$, $p = 0.53$). Most importantly, the interaction between condition and pictorial depth (Ferns: $F(2,10) = 4.553$, $p = 0.039$; Leaves: $F(2,10) = 4.493$, $p = 0.04$) and the interaction between condition and physical distance (Ferns: $F(2,10) = 8.526$, $p = 0.033$; Leaves: $F(2,10) = 23.54$, $p = 0.005$) were both significant. Neither the interaction between physical distance and pictorial depth (Ferns: $F(2,10) = 1.261$, $p = 0.325$; Leaves: $F(2,10) = 0.618$, $p = 0.559$) nor the interaction between condition, physical distance and pictorial depth (Ferns: $F(2,10) = 1.671$, $p = 0.237$; Leaves: $F(2,10) = 0.018$, $p = 0.982$) were significant.

4. DISCUSSION

In this study observers were asked to point in depth within a virtual pictorial space under both binocular- and monocular-aperture viewing. The positions in space to which participants pointed to differed considerably according to the viewing mode. Pointing responses under monocular viewing were mainly modulated by different pictorial depths. On the other hand, pointing responses under binocular viewing were mainly modulated by the changes in physical distance. The finding that the response patterns in the monocular condition were very similar for both physical distances is consistent with the fact that there were no reliable visual or non-visual signals indicating change in the distance of the image plane. The similarity in the responses can be attributed to default distance tendencies (e.g., specific distance tendency[23]). The remnant modulation of the monocular responses due to changes in physical distance may be ascribed to a residual effect of accommodation which is known to be available for distance estimation at these distances, although imprecise and inaccurate.[24] In contrast, information about the physical distance of the image plane in the binocular condition was provided by both vergence[25] and vertical disparities.[26] It is important to note, however, that, despite the richer distance information, pointing responses under binocular viewing did not correspond to the veridical physical distances. The range of responses was approximately half of the range of physical distances. This is consistent with the fact that depth estimation is rarely veridical.[27]

One possible explanation for the modulation of reaching responses under monocular viewing is that subjects were merely following a cognitive strategy of reaching based on the inferred relative positions of the targets within pictorial space. However, this explanation can be ruled out based on two observations. If the modulation was merely due to cognitive inferences, the response patterns should have been similar for both monocular and binocular viewing modes. Also, reaching responses should not have been very consistent over repeated and randomized presentations of the two images. Instead, individual variations over time under monocular viewing were minor; that is, observers appeared to conform to a consistent pictorial space. One concern might be that pointing in depth under binocular viewing is more unnatural. However, observers did not report a greater difficulty in this task compared to monocular viewing; they found the instructions under both modes of viewing very intuitive.

Given the small modulation of reaching responses attributable to pictorial depth in the binocular condition, an alternative explanation for the results is that they merely reflect the fact that a greater amount of relative pictorial depth between points was perceived under monocular compared to binocular viewing. This explanation is based on the fact that binocular viewing creates a cue-conflict situation (binocular disparity signals a flat surface in contradiction to the pictorial cues), and therefore, that a greater magnitude of relative depth will be perceived under monocular viewing due to reduction in cue conflict.[4,28] This explanation is, however, inconsistent with recent studies that clearly show little or no difference in estimates of relative depth between monocular and binocular viewing of pictorial images for a variety of tasks; these include depth estimates from curvature,[11] relative size,[29] dihedral angle[30] and slant.[31] On the other hand, the large individual differences in depth estimates inferred from the reaching responses under monocular viewing are consistent with large differences found in prior studies ("the beholder's share").[11,32,33]

The above analysis suggests that the difference in pointing responses between monocular and binocular viewing is not due to differences in the representation and perception of relative depth magnitude under the two viewing conditions, but rather, to differences in the representation of depth used to guide motor function. Programming reaching or pointing movements in depth requires absolute depth estimates; in other words, representation of absolute depth differences between points in arbitrary egocentric units. Pictorial depth cues themselves provide only relative depth information and need to be scaled by an estimate of viewing distance to generate scaled (absolute) depth estimates. The difference in reaching response between binocular and monocular-aperture viewing suggests differences in access to egocentric distance and scaled (absolute) depth representations between these two viewing conditions. It also provides supportive evidence that it is the availability of scaled depth representation that determines the degree of tangibility of visual space.[11]

REFERENCES

[1] Pirenne, M. H., [*Optics, Painting & Photography*], Cambridge University Press, Cambridge, UK (1970).

[2] Kubovy, M., [*The Psychology of Perspective and Renaissance Art*], Cambridge University Press, New York, NY (1986).

[3] Vishwanath, D., Girshick, A., and Banks, M. S., "Why pictures look right when viewed from the wrong place," *Nature* **8**, 1401–1410 (2005).

[4] Doorschot, P. C. A., Kappers, A. M. L., and Koenderink, J. J., "The combined influence of binocular disparity and shading on pictorial shape," *Percept. Psychophys.* **63**, 1038–1047 (2001).

[5] Wagemans, J., van Doorn, A. J., and Koenderink, J. J., "Pictorial depth probed through relative sizes," *i-Perception* **2**, 992–1013 (2011).

[6] Wheatstone, C., "Contributions to the physiology of vision, part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philos. Trans. R. Soc. Lond.* **128**, 371–394 (1838).

[7] Ames, A., "The illusion of depth in pictures," *J. Opt. Soc. Am.* **10**, 137–148 (1925).

[8] Schlosberg, H., "Stereoscopic depth from single pictures," *Am. J. Psychol.* **54**, 601–605 (1941).

[9] Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L., "On so called "paradoxical monocular stereoscopy"," *Perception* **23**, 583–594 (1994).

[10] Koenderink, J. J., "Pictorial relief," *Phil. Trans. R. Soc. Lond. A* **356**, 1071–1086 (1998).

[11] Vishwanath, D. and Hibbard, P. B., "Seeing in 3-D with just one eye: Stereopsis without binocular vision," *Psychol. Sci.* **24**, 1673–1685 (2013).

[12] Wade, N. J., Ono, H., and Lillakas, L., "Leonardo da Vinci's struggles with representations of reality," *Leonardo* **34**, 231–235 (2001).

[13] Wijntjes, M. W. A., Volcic, R., Pont, S. C., Koenderink, J. J., and Kappers, A. M. L., "Haptics disambiguates vision in the perception of pictorial relief," in [*Human Vision and Electronic Imaging XIV*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **7240**, 72400L (2009).

[14] Wijntjes, M. W. A., Volcic, R., Pont, S. C., Koenderink, J. J., and Kappers, A. M. L., "Haptic perception disambiguates visual perception of 3D shape," *Exp. Brain Res.* **193**, 639–644 (2009).

[15] Rogers, B. and Graham, M., "Similarities between motion parallax and stereopsis in human depth perception," *Vision Res.* **22**, 261–270 (1982).

[16] Ponce, C. R. and Born, R. T., "Stereopsis," *Curr. Biol.* **18**, R845–R850 (2008).

[17] Barry, S., [*Fixing my Gaze*], Basic Books, New York, NY (2009).

[18] Servos, P., Goodale, M. A., and Jakobson, L. S., "The role of binocular vision in prehension: a kinematic analysis," *Vision Res.* **32**, 1513–1521 (1992).

[19] Melmoth, D. R. and Grant, S., "Advantages of binocular vision for the control of reaching and grasping," *Exp. Brain Res.* **171**, 371–388 (2006).

[20] Nicolini, C., Fantoni, C., Mancuso, G., Volcic, R., and Domini, F., "A framework for the study of vision in active observers," in [*Human Vision and Electronic Imaging XIX*], Rogowitz, B. E., Pappas, T. N., and de Ridder, H., eds., *Proc. SPIE* **9014** (2014).

[21] R Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2013). http://www.R-project.org/.

[22] Wickham, H., [*ggplot2: Elegant Graphics for Data Analysis*], Springer, New York, NY (2009). http://had.co.nz/ggplot2/book.

[23] Gogel, W. C., "The sensing of retinal size," *Vision Res.* **9**, 3–24 (1969).

[24] Fisher, S. K. and Ciuffreda, K. J., "Accommodation and apparent distance," *Perception* **17**, 609–621 (1988).

[25] Foley, J. M., "Binocular distance perception," *Psychol. Rev.* **87**, 411–434 (1980).

[26] Rogers, B. J. and Bradshaw, M. F., "Vertical disparities, differential perspectives and binocular stereopsis," *Nature* **361**, 253–255 (1993).

[27] Volcic, R., Fantoni, C., Caudek, C., Assad, J. J., and Domini, F., "Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination," *J. Neurosci.* **33**, 17081–17088 (2013).

[28] Young, M. J., Landy, M. S., and Maloney, L. T., "A perturbation analysis of depth perception from combinations of texture and motion cues," *Vision Res.* **33**, 2685–2696 (1993).

[29] Wijntjes, M. W. A. and Pont, S. C., "Perceived depth in photographs: humans perform close to veridical on a relative size task," *J. Vis.* **12**(9), 277 (2012).

[30] Cooper, E. A. and Banks, M. S., "Perception of depth in pictures when viewed from the wrong distance," *J. Vis.* **12**(9), 896 (2012).

[31] Erkelens, C. J., "Virtual slant explains perceived slant, distortion, and motion in pictorial scenes," *Perception* **42**, 253–270 (2013).

[32] Koenderink, J. J. and van Doorn, A. J., "Relief: pictorial and otherwise," *Image Vision Comput.* **13**, 321–334 (1995).

[33] Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., and Todd, J. T., "Ambiguity and the "mental eye" in pictorial relief," *Perception* **30**, 431–448 (2001).