Different cue weights at the same place

C. M. P. Muller Research Institute MOVE, Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

E. Brenner Research Institute MOVE, Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

J. B. J. Smeets Research Institute MOVE, Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

The visual system uses multiple cues to estimate properties of interest. Since the errors in the estimates from different cues for the same property are generally different, a weighted average of the cues provides a better overall estimate. The most precise estimate is found when each cue's weight is proportional to its reliability. We here show that the weights given to cues for surface slant can differ between two transparent surfaces that are at the same location at the same time. Thus the weights must be assigned separately for each structure, rather than for each location.

Keywords: 3D surface and shape perception, binocular vision, spatial vision, space and scene perception

Citation: Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2009). Different cue weights at the same place. *Journal of Vision, 9*(11):26, 1–5, http://journalofvision.org/9/11/26/, doi:10.1167/9.11.26.

Introduction

The visual system uses multiple complementary sources of information (cues) to estimate properties of interest. Since the errors in the estimates from different cues for the same property will generally be different, a weighted average of the cues provides a better overall estimate. The most precise estimate is found when each cue's weight is proportional to its reliability (Backus & Banks, 1999; Ernst & Bülthoff, 2004; Hillis, Watt, Landy, & Banks, 2004; Landy, Maloney, Johnston, & Young, 1995; van Beers, Sittig, & Denier van der Gon, 1996). But how is this reliability known? Is it based on experience or on the information in the image at that moment? Is it determined for regions of a scene or for separate items in the scene?

What if we want to judge the slant of the surface of a textured rectangular table with a ring left on it by a glass of wine from which a bit had been spilt the previous evening? In the images reaching our eyes, the outline of the table's surface, the shape of the ring, the texture gradient, and the gradient in binocular disparities all provide information about the surface's slant. When considering the whole surface, the ring will contribute to the binocular disparity gradients, it may slightly disrupt the texture cue, and the shape of its image will provide independent information about the slant. If we are sure that all the cues, including the shape of the ring, relate to a single surface with a single slant, we can best estimate that slant by combining all the cues. The absence of discontinuities in the texture and disparity gradients may justify assuming that there is a single surface with a single slant. Combining all the cues to estimate the surface's slant means that one may end up with a different judgment of the orientation of the ring when considering it as part of the surface than one would if one were to judge its orientation independently.

It is well established that the weights given to different cues can depend on the task (e.g. Bradshaw, Parton, & Glennerster, 2000; Glennerster, Rogers, & Bradshaw, 1996; Koenderink, Kappers, Todd, Norman, & Phillips, 1996; Tittle, Norman, Perotti, & Phillips, 1998) but it is not evident that judging the slants of the ring and the table are fundamentally different tasks. Neither is it clear whether cues' weights can differ for different structures within confined regions of the visual field, because in order to do so the structures first have to be segregated. On the other hand, if the same slant cue weights are assigned for all structures within some region of space, then these weights cannot be optimized for both the ring and the surface texture.

Here, we examine how binocular and monocular cues are combined for the perception of surface slant. The reliability of slant cues depends on many factors, such as the slant angle, the viewing distance and the structure of the image (Jacobs, 2002; Knill, 1998; Muller, Brenner, & Smeets, 2007). There is some evidence that information about the reliability under the prevailing conditions is learnt from experience (Jacobs & Fine, 1999; Knill, 2007), although the reliability could also be estimated from the properties of the images at each moment (Deneve, Latham, & Pouget, 2001). In either case the reliability could be estimated for regions in space or for items within

doi: 10.1167/9.11.26 Received July 31, 2009; published October 28, 2009 ISSN 1534-7362 © ARVO

that space. In the present study we attempt to shed some light on the framework within which slant cue weights are attributed.

Methods

Rationale

If an estimated slant is a weighted average of binocular and monocular cues, we can determine the weight given to each by asking observers to estimate the slant of a cue conflict surface in which monocular and binocular cues indicate different slants (Louw, Smeets, & Brenner, 2007; Young, Landy, & Maloney, 1993). To distinguish between assigning weights to cues within a certain region and assigning weights to cues within each item, we used two simulated transparent cue conflict surfaces that were at the same place. One surface was a sparsely dotted plane, which provides reliable information from binocular disparity. Monocular information about slant, for instance obtained from the texture gradient (assuming an isotropic texture), was not very reliable due to the low density and random distribution of the dots. The other surface was a ring, the slant of which could be judged reasonably reliably from the aspect ratio in each monocular image (assuming that the ring is a true circle). Thus if cue reliability is estimated independently for each surface, we expect monocular and binocular cues to be assigned different weights for the two surfaces. If it is estimated for a part of the visual field, we expect cues to be assigned the same weights for the two surfaces, and the surfaces to be perceived as having the same slant when the slants specified by the cues are the same.

Setup

Our setup consisted of an Apple G5 computer that generated the images and registered the responses, a 57 cm (diagonal) Sony Trinitron monitor for presenting the images (resolution 1096 × 686 pixels), and Crystal Eyes stereo shutter spectacles to present a different image to each eye. The images were generated at a refresh rate of 160 Hz (80 Hz per eye), using only the red gun because the spectacles work best for red images. Observers sat 1 meter from the screen, so that the screen was approximately 27° × 17°.

Stimuli

Each stimulus consisted of two surfaces (Figure 1). One surface was a disc defined by 25 randomly distributed red dots. Either all slant cues indicated that this disc of dots had a 45° slant (consistent), or else binocular disparity and the monocular cues were in conflict. When the cues were in conflict, either the binocular cue indicated that the slant was 37.5° and the monocular cues that it was 52.5° (conflict 1, as depicted in the side view), or vice versa (conflict 2). Each of the three conditions (consistent; conflict 1; conflict 2) was presented 30 times, in a randomly interleaved order. The other surface was a ring

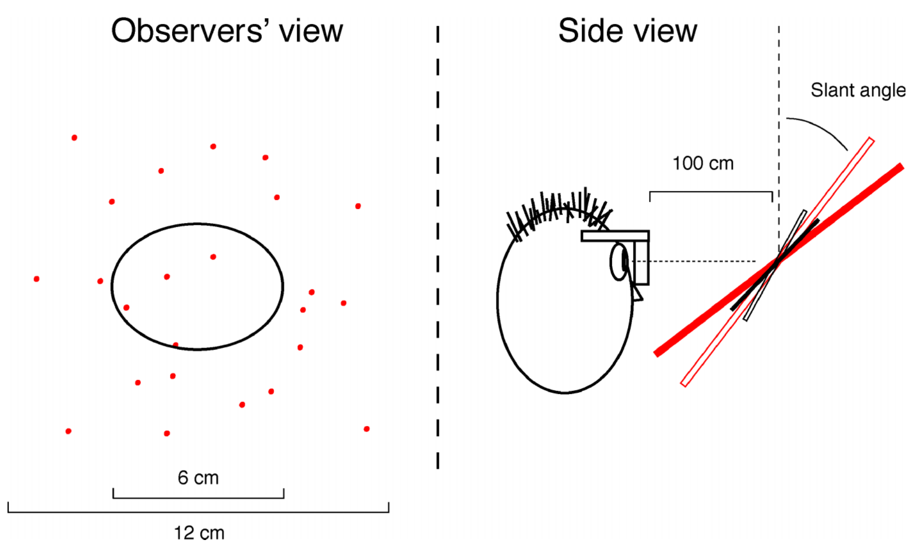

Figure 1. Observers viewed a 3D simulation of two transparent surfaces slanted around a common horizontal axis. Both the ring and the dots were red in the experiment, but the ring is depicted in black here for clarity. Open bars show the slant indicated by binocular disparity. Solid bars show the slant indicated by monocular cues. In this example, the slant indicated by the monocular cues was 15° larger than that indicated by binocular disparity.

that had the same 0°, 15° or −15° conflict between the cues as the disc of dots (on every trial), but the observer could vary its slant. When the observer did so, the two cues changed in synchrony, so that the cue conflict (or lack thereof) did not change.

The ring was approximately 3.5° wide. The disc of dots was approximately 7° wide. The information about slant was in the global image shape and in the gradients of texture density and binocular disparity. In order to judge either of these surfaces' slants observers therefore had to take an extended region of the visual scene into account.

Procedure

Nine observers participated in three sessions. In two sessions they performed a matching task and in the third a decision task. In the matching task observers set the slant of the ring to match that of the dotted plane by moving the computer mouse. Moving the mouse to the left decreased the slant, and moving it to the right increased the slant. Once satisfied with their match, observers clicked the mouse button to start the next trial. There was no time limit for making the settings. When setting the slant, the values for both the monocular and binocular cues for slant were adjusted, but the discrepancy between them was kept constant. In the matching task the observers could see that the ring was not part of the disc of dots by the motion of the ring when they changed its slant. In the decision task, which was performed to determine whether the motion of the ring was critical for performing the task, the same observers were presented with a static display of the ring and the dotted plane for 1500 ms. The ring had one of 7 possible slants (2° steps, centered on the veridical match, 30 trials each). Observers pressed a key to indicate whether they perceived the ring to be more or less slanted than the dotted plane.

In the session with the decision task and in one of the sessions with the matching task the conditions were as described above. The purpose of the second matching session was to estimate the weights attributed to the binocular and monocular cues for each of the two different surfaces. In this session there were four conditions. Within each condition, either the ring or the disc of dots was identical to that in one of the two conflict conditions of the initial matching session. The other was identical to that in the consistent condition. Thus on each trial there was one conflict and one non-conflict surface.

Analysis

The slants set in the matching task were averaged per observer and per condition. We tested whether the differences between the average settings for the two surfaces were consistent across observers using *t*-tests (for each condition). For the decision task settings, points of subjective equality were determined by fitting psychometric curves to the fraction of "more slanted" responses as a function of the slant of the ring (for each condition and each observer). To estimate the weights from the second matching session we compared the value of the slant with consistent cues with the two values of the slants with inconsistent cues. By doing so we obtained an estimate of the weights for each cue conflict for each surface. With these weights we estimated the perceived slants of the ring and the disc of dots for the cue values set in the main experiment.

Results

In the consistent condition of the matching task, observers aligned the surfaces very accurately ($t(8) = 0.60$, $p = 0.56$; see Figure 2). In the cue conflict conditions they made systematic errors ($t(8) = 4.15$, $p = 0.003$ and $t(8) = -2.39$, $p = 0.04$). They could have matched the cues (both open and closed symbols) across the surfaces, but did not do so. The average error (i.e. the difference between the red squares and black circles) was 2.4°. The decision task data shows a very similar pattern ($t(8) = 5.98$, $p < 0.001$ and $t(8) = -3.29$, $p = 0.01$ for the conflict conditions; $t(8) = 1.67$, $p = 0.13$ for the consistent condition), except that the ring had to be slightly more slanted than the disc of dots to appear to be aligned with it: in Figure 2 the closed black symbols are 0.9° higher for the decision task than for the matching task. This bias was present in both the conflict conditions and the consistent condition.

The cue weights for each of the two surfaces were estimated from the average matches in the second matching session, in which the same observers matched cue consistent stimuli to cue conflict ones. The estimate of the weight given to the binocular cue was 71% for the ring and 88% for the disc of dots. These values were used to calculate hypothetical perceived orientations for the settings in the first two sessions. The hypothetical perceived orientations are represented by black and red crosses in Figure 2. These orientations were quite similar for the two (matched) surfaces; much more so than the values of the individual cues. The difference between the average binocular weights for the two kinds of surfaces was 18%. For this difference, and a 15° conflict, the individual cues can be expected to differ in slant by 2.7° for the orientations to look the same. As already mentioned, the average measured difference was 2.4°, which is close to this prediction. Thus it would appear that observers matched the differently weighted averages rather than the individual cue values.

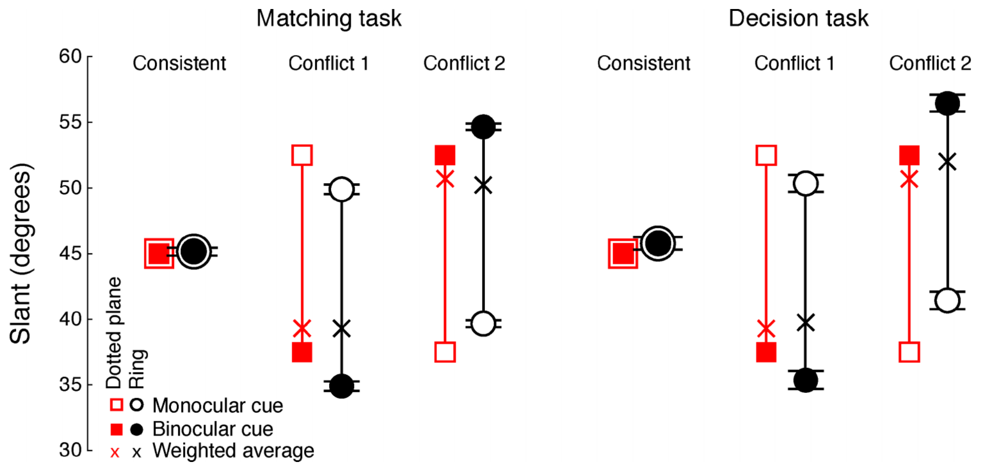

Figure 2. Cue values for which the ring and the dotted plane appear to have the same slant. Observers' average settings in the matching task are plotted on the left. Points of subjective equality from the decision task are plotted on the right. Horizontal lines indicate between-observer standard errors. The crosses mark estimated weighted averages for perceived slant based on a separate session in which cue consistent and cue conflict surfaces were matched.

Discussion

In the main sessions, the two surfaces had the same cue conflict. Observers could have made settings that aligned the values of all the cues; they could have set the circles and squares to the same value in Figure 2 (thus avoiding discontinuities). However, they did not. Observers matched a weighted average instead. This means that the weights given to cues for surface slant are assigned separately for each surface, and can differ between two different types of transparent surfaces although they are at the same location at the same time. Observers assign the weights in accordance with the cue precisions within each separate surface, and match the different weighted averages. A consequence of this is that if the two surfaces have exactly the same values for each of the cues, the weighted averages will differ, so observers will perceive them as having different slants.

It is important to realize that we are not distinguishing between the weights given to the two cues, but between their weights for different structures (at the same place). Of course for this the structures have to be considered to potentially have different slants, so that the cues from the different structures are not averaged to get a better overall estimate of slant. This is not trivial, because different cues for the slant of a single surface will often depend to a different extent on different parts of the image. Thus one will often need to combine judgments across the surface.

In the matching task it is evident from the motion of the ring that occurs when it is being adjusted that the two surfaces have independent slants. Since the results of the decision task are very similar, apart from a small bias that is present in all three conditions of the decision task, the motion of the ring in the matching task is apparently not critical for treating the two surfaces differently. This follows already from the fact that observers could perform the task; if they would have considered the two surfaces to form a single surface with a single slant, they would not have been able to perform our experiment which obviously required observers to consider them as two separate surfaces.

Apparently, at least when asked to treat two surfaces separately, cue weights are assigned separately for each surface, even if they are in the same spatial region. A possible explanation for this is that the reliability is judged together with the slant itself, on the basis of the same information. For instance, the reliability of the judged slant from the shape of the retinal image of the ring could be judged from the resolution of the judgments of the separation between the borders in different directions, which in turn could be judged from how precisely one can localize each border. The latter could be judged from the spread of activation across cells with slightly different spatial fields within the brain (Knill & Pouget, 2004). Such estimates could be combined with biases from previous experience (Knill, 2007) in order to find the best estimate for each structure in the scene. This way of assigning cue weights is a good strategy when different cues provide reliable information for different structures on a single surface. In our study, such different structures were on separate surfaces, but presumably, if observers had been asked to judge the slant of a single dotted surface with a ring on it, they would also assign these weights to each cue within each structure and then combine the slants indicated by the two structures. Doing so would give the best overall estimate. The most striking thing, though, is that even when the surfaces are considered separately, we find no evidence at all for comparing the surfaces at the level of the individual cues.

Acknowledgments

This research was funded by the Dutch Science Foundation (NWO, Grant # 452-02-007).

Commercial relationships: none.
Corresponding author: Chris Muller.
Email: christiaan.muller@ugent.be.
Address: Research Institute MOVE, Faculty of Human Movement Sciences, Vrije Universiteit, Van der Boechorststraat 9, 1081 BT Amsterdam, The Netherlands.

References

Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception, 28,* 217–242. [PubMed]

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research, 40,* 3725–3734. [PubMed]

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4,* 826–831. [PubMed]

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Science, 8,* 162–169. [PubMed]

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36,* 3441–3456. [PubMed]

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12):1, 967–992, http://journalofvision.org/4/12/1/, doi:10.1167/4.12.1. [PubMed] [Article]

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Science, 6,* 345–350. [PubMed]

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research, 39,* 4062–4075. [PubMed]

Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research, 38,* 1683–1711. [PubMed]

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision, 7*(8):13, 1–20, http://journalofvision.org/7/8/13/, doi:10.1167/7.8.13. [PubMed] [Article]

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neuroscience, 27,* 712–719. [PubMed]

Koenderink, J. J., Kappers, A. M., Todd, J. T., Norman, J. F., & Phillips, F. (1996). Surface range and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 869–878. [PubMed]

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Louw, S., Smeets, J. B., & Brenner, E. (2007). Judging surface slant for placing objects: A role for motion parallax. *Experimental Brain Research, 183,* 149–158. [PubMed] [Article]

Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2007). Living up to optimal expectations. *Journal of Vision, 7*(3):2, 1–10, http://journalofvision.org/7/3/2/, doi:10.1167/7.3.2. [PubMed] [Article]

Tittle, J. S., Norman, J. F., Perotti, V. J., & Phillips, F. (1998). The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture, and shading. *Perception, 27,* 147–166. [PubMed]

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research, 111,* 253–261. [PubMed]

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