Is prior knowledge of object geometry used in visually guided reaching?

Bruce Hartung	Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA	\bowtie
Paul R. Schrater	Departments of Computer Science & Engineering and Psychology, University of Minnesota, Minneapolis, MN, USA	
Heinrich H. Bülthoff	Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany	
Daniel Kersten	Department of Psychology, University of Minnesota, Minneapolis, MN, USA	
Volker H. Franz	Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany, & University of Giessen, Giessen, Germany	

We investigated whether humans use prior knowledge of the geometry of faces in visually guided reaching. When viewing the inside of a mask of a face, the mask is often perceived as being a normal (convex) face, instead of the veridical, hollow (concave) shape. In this "hollow-face illusion," prior knowledge of the shape of faces dominates perception, even when in conflict with information from binocular disparity. Computer images of normal and hollow faces were presented, such that depth information from binocular disparity was consistent or in conflict with prior knowledge of the geometry. Participants reached to touch either the nose or cheek of the faces or gave verbal estimates of the corresponding distances. We found that reaching to touch was dominated by prior knowledge of face geometry. However, hollow faces were estimated to be flatter than normal faces. This suggests that the visual system combines binocular disparity and prior assumptions, rather than completely discounting one or the other. When comparing the magnitude of the hollow-face illusion in reaching and verbal tasks, we found that the flattening effect of the illusion was similar for verbal and reaching tasks.

Keywords: visual motor control, reaching, haptic feedback, hollow-face illusion

Introduction

Current literature on both robotic and human reaching assumes that most of the information used when planning and executing a visually guided reach is visually available at the time of the reach. In other words, it is assumed that prior knowledge is not used. The exception to this is prior knowledge about calibrations, such as camera calibrations and calibrations between cameras and manipulators or between the eye and the hand for human observers. However, given the well-known ambiguities in visually extracting object shape (Belheumer, Kriegman, & Yuille, 1999), the use of prior information for shape may be critical for making successful visually guided reaches. In this study we explore whether or not visually guided reaching in humans uses prior knowledge of the geometry-specifically, prior knowledge of the geometry of faces.

Faces are convenient targets for our experiments, in part because of the well-known "hollow-face illusion" (Gregory, 1973). When an observer views the inside of a

mask or mold of a face, depth estimates from binocular disparity conflict with prior knowledge of the shape of faces. It seems that prior knowledge "wins," and the mask is seen as a convex face (i.e., having normal geometry). This is the "hollow-face illusion."

One may wonder if the hollow-face illusion is simply a manifestation of a general convexity bias, as demonstrated by Langer and Bülthoff (2001), rather than depending on familiarity with faces. If this were the case, we would expect that inverted versions of less familiar objects would exhibit the same effect as the hollow-face illusion, and that the effect of the illusion would be of the same magnitude. However, this is not the case. As shown by Hill and Bruce (1994), a "hollow-potato illusion" has a smaller effect on verbal tasks than does the hollow-face illusion. This suggests that the hollow-face illusion is more than just a manifestation of a general convexity bias and that prior knowledge of an object's geometry is used when making verbal depth estimates. However, it does not immediately follow that the visual system guides reaches using the same prior knowledge. The questions of how prior knowledge and binocular disparity are combined and what strategies are used to combine these cues are also still open.

In this study, we use the hollow-face illusion to test whether prior knowledge of geometry is used when making visually guided reaches, and whether this knowledge combines with or supercedes binocular stereo information in both reaching and verbal tasks. Finally, we investigate whether the use of prior knowledge depends on the task performed by comparing shape estimates from verbal tasks to those from reaching tasks. We will briefly discuss each of these research questions.

1.1 Does prior knowledge affect reaches to faces?

In theory, reaches could be controlled completely by information present at the time of the reach, for example, binocular disparity (Hespanha, Dodds, Hager, & Morse, 1999). On the other hand, prior knowledge of an object's geometry could be used in combination with binocular disparity to make a more accurate estimate of the object's geometry. If the visual system does use prior knowledge of an object's geometry, then reaches should be affected by the hollow-face illusion, as tested by our experiment.

Because previous work (Hill & Bruce, 1993, 1996) has shown that the hollow-face illusion affects verbal estimates of face geometry, one may be tempted to assume that other types of tasks will also be affected. However, Schrater and Kersten (2000) have shown that optimal cue combination depends on the task being performed. For us, this may mean that the optimal combination of prior knowledge and binocular disparity is different for verbal tasks than for reaching tasks. Prior knowledge may be a cue that dominates when the task is to verbally estimate a familiar object's geometry, whereas binocular disparity may dominate when the task is to guide reach to the same object. Indeed, some studies (e.g., Bridgeman, Lewis, Heit, & Nagle, 1979; Bridgeman, Peery, & Anand, 1997; Milner & Goodale, 1995) suggest that illusions that affect some verbal (or more general "perceptual") tasks, do not affect the visual control of reaching tasks and that different types of reaching tasks may be affected differently. More specifically, if haptic feedback is provided at the end of the reach, the reach may not be affected by the illusion, but if haptic feedback is not provided, it will. However, target stimuli used in the cited experiments were not designed to test for the effects of prior knowledge. In those experiments, participants estimated size, length, or position of abstract geometric entities, such as lines and circles. As these do not have a "typical" or expected size, these experiments did not test the effects of prior knowledge inherent in the hollow-face illusion used in our experiment.

While the previous section asks if the visual system uses prior knowledge, it does not ask how information from prior knowledge of geometry is combined with information from binocular disparity. When prior knowledge and disparity information are in conflict, it is possible that the visual system uses a winner-take-all strategy-using only prior knowledge when reaching to sufficiently familiar objects. It is also possible that the visual system combines these sources of information to yield a shape estimate that forms a compromise between the two sources. For example, the visual system may use a weighted combination of the information from binocular disparity and prior knowledge of geometry. We will compare reaches to hollow faces with reaches to normal faces. If the visual system uses a winnertake-all strategy in which prior knowledge of geometry is the winner, reaches should be the same for both hollow and normal faces.

1.3 Is the magnitude of the hollow-face illusion task dependent?

While the first question asks whether prior knowledge is used for reaching tasks, it does not ask whether prior knowledge affects the three types of tasks equally. We will compare the magnitude of the effect of the illusion on all three tasks.

In our experiment, we presented participants with computer-generated images of convex (normal) and concave (hollow) faces, such that depth information from binocular disparity was consistent or in conflict with prior knowledge of the geometry. We used an experimental setup that enabled us to minimize other potential cues for face geometry. For example, faces were rendered as Lambertian surfaces with directional light sources, such that shading would not bias the participants to the concavity or convexity of the faces. Participants reached to either the nose or cheek of the faces or gave verbal estimates of the corresponding distances. If prior knowledge about the geometry of faces does affect participants' reaches, we expect them to reach to concave faces as if they were convex; therefore, we expect them to reach to the nose as if it were in front of the cheek, even though it is behind.

Methods

Participants viewed concave and convex faces and made verbal and reach estimates that indicated the participant's perceived shape of the face.

2.1 Participants

Five naïve University of Tübingen students took part in the study. In return for their participation, they received a payment of 13 DM (app. 6.5 US\$ or 6.5 EURO) per hour.

2.2 Apparatus

The faces were stereo pairs rendered using OpenGL, scaled to normal size for an adult head. The faces were taken from the Tübingen Face Database (http://faces.kyb.tuebingen.mpg.de; Troje & Bülthoff, 1996; Blanz & Vetter, 1999). For the sake of simplicity, it was important to choose a lighting model that would not add an additional source of information for determining the concavity/convexity of the face. To that end, each face was rendered as a Lambertian surface, lit by a single, directional light source along the view direction. Because the concavity/convexity of the face was determined by a scaling in the view direction, this light source created shading that provided only ambiguous information for determining the concavity or convexity of the face.

It was necessary to present the faces in such a way that participants would be able to reach to the perceived location of the face images. To achieve this, stimuli were rendered on a CRT suspended above a mirror, as shown in Figure 1. The faces' location as defined by binocular disparity and perspective cues was behind the mirror. As shown in the figure, this setup allowed participants to place their hands at the location of the faces.

A chin rest and headrest were used to maintain a consistent viewing position. For the reaching tasks, the participant's right index finger was placed into the thimble of a PHANToMTM force feedback device that was used to give haptic feedback as well as measure the trajectory of the finger.

2.3 Procedure

Each participant performed one verbal and two reaching tasks. Before the trials began, participants were instructed about which parts of the nose and cheek were the targets. So that participants would not be biased, they were not told to touch the face from the inside or from the outside, just to approach the target from "the side." In each task, one of three faces was presented at a distance of 460, 490, or 520 mm from the viewpoint to the center of the face. This range was selected because the stereo-graphics effect began to degrade for faces closer than 460 mm, and participants were not able to reach faces further than 520 mm, due the configuration of the haptic workspace. Faces were presented in two different orientations. In half of the trials, the faces were oriented such that the participant viewed a normal (convex) face, and in the other half, they viewed a hollow (concave) face. The 36 possible trial types (3 faces x 3 distances x 2 targets x 2 orientations) were randomized within trials. The randomization and the relatively large number of possible trial types make it unlikely that participants were able to guess in which condition they were. In concave trials, the nose was at the same distance from the viewer as the cheek was during convex trials. Likewise, in concave trials, the cheek was at the same distance from the viewer as the nose was during convex trials.



Figure 1. Participants viewed computer-generated images of a face in stereo. The image was reflected from a CRT onto a mirror. Participants were able to interact with the graphics at the location of the image, underneath the mirror. Haptic feedback was provided by a PHANTOM[™] force feedback device. (Adapted from an illustration by Marc O. Ernst.)

The order of tasks (verbal, haptic, and non-haptic) was randomized within participants.

In the verbal task, participants were asked to give a verbal estimate of the distance from their viewing position to either the nose or the cheek of the faces. Estimates were given in arbitrary units, chosen by the participant. The participants were instructed that their eyes were at zero, and were told to use any metric they were comfortable with, so long as they were consistent. In each trial, the face was shown and a tone sounded. The face was removed from view after 2 s and a second tone sounded. Participants were instructed to respond before the second tone. This limit was imposed to keep the response time similar between the reaching and verbal estimates. Each participant made distance estimates for two types of targets (nose or cheek) on two types of faces (concave and convex) at three distances (460, 490, and 520 mm). Each condition was repeated 33 times for a total of 396 trials in the verbal task.

In the non-haptic reaching task, participants were asked to touch either the nose or cheek of the face. The mirror occluded the participant's finger, but a "virtual finger" in the form of a ball was presented at the position of the fingertip. The finger was not visible at its starting point. In each trial, a face was shown and a tone sounded. The face was removed as soon as the finger came into view. Because the face and finger were both rendered objects, we were able to ensure that the finger and face were never visible at



Figure 2. a. Faces were rendered in graphics space. Participants were asked to reach and touch the side of the nose or the cheek of the face. b. During haptic trials, a "board" was rendered in haptic space to give feedback at the correct *X* and *Y*, but at any *Z*. c. This figure shows the two spaces in relation to each other. Note that participants could not see the haptically rendered board.

the same time. A second tone sounded 2 s after the first. Participants were instructed to complete the reach before the second tone. This limit was imposed to keep the response time similar between the reaching and verbal estimates. The final Z-position of the finger was recorded as the estimated depth of the target, where the Z-axis is the view direction, with its origin between the participant's eyes. Participants were asked to touch the side of the nose, or the side of the cheek, so this reach would be consistent with the haptic task, described below. Each participant made reaches to exactly the same stimulus conditions as in the verbal tasks (to a total of 396 trials).

The haptic task was similar to the non-haptic task, with the addition of haptic feedback at the tip of the index finger. In all other respects, the haptic task was identical to the non-haptic task (again, participants performed a total of 396 trials). To ensure that the haptic feedback did not give information about the true distance to the target, ambiguous haptic feedback was given. As shown in Figure 2, a board was rendered in haptic space using a PHANToMTM force-feedback device. The position of the board in the X direction was consistent with the X position of the target nose or cheek in the trial. The board was fairly short (4 cm) in the Y direction, so participants would miss the board if their reaches were not accurate in the Y direction. The board gave the correct feedback at the X and Y coordinates of the target (nose or cheek), but at any Z (distance). In this way we (a) ensured that participants received haptic feedback. This is important because lack of haptic feedback might change the planning and dynamics of the reaching movement (e.g., Goodale, Jakobson, & Keillo, 1994); (b) we excluded the possibility that participants adopted a strategy to simply move the finger forward until they touched the target object. In this case, participants would

not need to exclusively use visual information, such that we could not draw inferences about the underlying visual information processing. The board was haptically rendered to be somewhat sticky, such that it discouraged the participants from sliding their fingers forward along the board, and discovering its true shape. In the post-experiment interview, participants were asked if they discovered anything strange about the shape of the face from touch. None reported that they did. They were asked directly if the haptic feedback was convincing, and all agreed that it was.

To be sure that the participants had not consciously used a different strategy during cue-conflict trials, they were interviewed after all trials were completed. Participants were first asked if they noticed anything different about any of the faces. Some participants noted that different face models were used, and that some were further away than others, but none reported noticing the inversion of the faces. Participants were then asked directly if they had noticed that some of the faces were concave. Again, none reported knowing that they were concave.

2.4 Data analysis

Each trial resulted in a measurement of the estimated distance to the nose or cheek. In the two reaching tasks, these distances were measured in millimeters, such that we could use these values directly for our analyses. In the verbal estimation task, estimates were given by the participants in arbitrary units. Therefore, we performed a normalization relative to the maximum estimate given by each participant, such that we calculated the estimated distance as percentage of this maximum estimate. Further, because we were interested in the perceived depth (thickness) of the face, not the distance to individual targets, we calculated for each orientation and depth, the average difference between the nose and cheek responses as a measure of the perceived

Results

We calculated repeated measure ANOVAs for the reached distance in the reaching tasks (2 task x 3 distance x 2 concave/convex x 2 nose/cheek) and for the estimated distance in the verbal estimation task (3 distance x 2 concave/convex x 2 nose/cheek). The results are also graphically depicted in Figure 4. We first describe the results of the ANOVAs in a compact fashion and then relate them to our research questions.

In the reaching task, we found a main effect of distance, F(2,8) = 4.8, p = .043, indicating that participants responded to larger distances with longer reaches (see upper panel of Figure 4). This effect was similar for the nonhaptic and the haptic tasks (main effect task was not significant: F(1,4) = 2.8, p = .171). Participants reached further to cheeks than to noses (main effect nose/cheek: F(1,4) = 10.3, p=.033). This effect was modulated by the hollow-face illusion (interaction nose/cheek x con-



Figure 3. The difference between the nose and cheek distance estimates was calculated as a measure of the participant's depth estimate.

cave/convex: F(1,4) = 8.3, p = .045) and was slightly different for the two tasks at different distances (interaction nose/cheek x distance x task: F(2,8) = 9.8, p = .007). All other main effects or interactions were not significant.



Figure 4. Upper row. Average distances estimated in the verbal task and reached in the two reaching tasks as function of the distance of the face, the target (nose vs. cheek), and the type of face (concave vs. convex). Lower row. Average depth (i.e., distance to the cheek minus distance to the nose) as a function of the type of face (concave vs. convex) for each of the three tasks. Error bars indicate ± 1 SEM. In the upper row we did not present error bars because here the SEM contain between and within participants variance and are therefore not informative for our repeated measures analysis.

In the verbal estimation task, we also found a main effect of distance, F(2,8) = 7.3, p = .016, indicating that participants responded to larger distances with larger estimates. Participants estimated cheeks further than noses (main effect nose/cheek: F(1,4) = 25.7, p = .007) and this effect was modulated by the hollow-face illusion (interaction nose/cheek x concave/convex: F(1,4) = 49.5, p = .002). All other main effects or interactions were not significant.

3.1 Does prior knowledge affect reaches to faces?

The first question that we wanted to answer was whether all three tasks were affected by the hollow-face illusion. If so, then the distance to the nose should be estimated to be less than the distance to the cheek for the concave faces, even though the distance to the nose was greater, as defined by binocular disparity. The depths (as the difference between the cheek and nose distance estimates) are plotted in the lower row of Figure 4. If participants respond veridically, the depths should be positive for convex faces, as the nose is closer to the observer, and negative for the concave faces and have the same magnitude as the convex estimate.

Inspecting the figure shows that in all tasks, convex as well as concave depth estimates were positive. This indicates that the hollow-face illusion did affect all three tasks (cf. the significant nose/cheek main effects in the ANOVAs). In all three tasks the depth estimates were decreased in the concave conditions relative to the convex conditions (cf. the nose/cheek x concave/convex interaction in the ANOVAs). This indicates that the binocular information is not totally discounted. But even if we calculate separate analyses for the concave conditions alone, we still get significantly positive effects in the reaching tasks, t(4) = 2.8, p = .048, and a strong trend in the verbal estimation task, t(4) = 2.7, p = .055. Both results together indicate that prior knowledge is stronger than binocular information, but cannot totally overwrite the binocular information.

Note that the depth-reducing effect of binocular information is similar for all three tasks (verbal estimation: concave depth is 44% of convex depth; non-haptic: 46%; haptic: 61%), which suggests similar cue combination strategies for the different tasks. In the following two sections we will further explore the questions of cue combination strategies by using the individual data from each participant (instead of the averaged group data). By this approach we are able to further justify our claims, because we exclude the possibility of artifacts that might be caused by averaging the data of single participants into group data.

3.2 How is prior knowledge combined with binocular disparity?

When viewing the concave faces, prior knowledge is in conflict with binocular disparity. We were interested in

how this conflict was reconciled. The visual system may make a weighted combination of disparity and priors, or it may use a winner-take-all strategy in which one is completely disregarded.

For each task, we plotted the depth estimate in the concave condition versus the depth estimate in the convex condition for each observer. The plots are shown in Figure 5. Each data point is the average for one participant. If prior knowledge completely dominates the depth estimates, the concave and convex depths should be the same, and we would expect the data points to lie on the line of slope 1.0 (which is plotted in red). If stereo information completely dominates the depth estimates, then we would expect the data points to lie on the line of slope -1.0 (which is plotted in blue). If the two cues are weighted, but prior knowledge is weighted more heavily, then the data points will lie above y = 0 (in the yellow wedge). Similarly, if stereo information is weighted more heavily, the data points will lie below y = 0 (in the green wedge). As can be seen, for all tasks and for all participants, the data points lie in the yellow wedge, and some of them even lie on the red slope=1 line. That is, prior knowledge dominates the depth estimates for all tasks for all participants. For participant ST, prior knowledge completely dominates the depth estimates in both reaching tasks.

While prior knowledge dominates the depth judgments, the presence of conflicting binocular disparity flattens the face for all participants in the verbal task, and for all but one participant in the reaching tasks. This means that although prior knowledge is very strong, it does not appear that the cue conflict is resolved with a winner-takeall strategy.

3.3 A comparison of the illusion's effect on each task

We found that the hollow-face illusion affected each of the three types of tasks. It is also of interest to compare the effects quantitatively. This is not straightforward when the comparison is between the verbal task and either of the two reaching tasks. The estimates given in the verbal task were in arbitrary units, chosen by each participant. The estimates given in the reaching tasks were in millimeters. To compare relative differences between these measures, we used the following geometrical analysis: Each data point in Figure 6 is the average depth estimate for one task, plotted against the average estimate for another task, for one orientation (concave vs. convex) for one participant. For example, consider the comparison of the non-haptic and verbal depth estimates for participants SH in Figure 6. The non-haptic depth estimate in the concave condition was half the size of the estimate in the convex condition. If the illusion had a similar effect on the perceived depth in the verbal task other than a change of units, we would expect a similar relationship in the reaching task. This is what we found. The verbal estimates of the participant SH in the concave





Figure 5. Average convex depth estimate versus the average concave depth estimate for each of the three tasks for each participant. For all tasks and for all participants, the data points lie above y = 0 (in the yellow wedge), and some of them lie on the (red) slope=1 line. That is, prior knowledge dominates the depth estimates for all tasks for all participants. For participant ST, prior knowledge completely dominates the depth estimates in both reaching tasks. Error bars indicate ±1 SEM.

condition were roughly half the size of the verbal estimates in the convex condition. More generally, if the data point of the concave condition lies on the line that connects the origin with the data point of the convex condition, then the illusion's effect on cue weighting is the same for both tasks. If the data of the concave condition lie above this line, then the illusion's effect is stronger for the task on the *y*-axis. If the data point of the concave condition lies below this line, then the illusion's effect is stronger for the task on the *x*-axis.

As shown in Figure 6, the illusion's effect is similar for three of the five participants when comparing the verbal task to either reaching task. One of the remaining participants shows a lesser effect on the reaching tasks (AL, shown in blue), and one shows a greater effect (ST, shown in green). That is, the weighting given to binocular disparity versus prior knowledge is the same for three of the five participants. For one participant, binocular disparity is weighted more heavily. For another participant, prior knowledge is weighted more heavily.

Discussion

4.1 Prior knowledge and reach

The first question we wanted to answer was whether the motor system uses prior knowledge about the objects that it is reaching to. We found that participants do not reach to a nose that is behind a cheek, that is, the motor system is affected by the hollow-face illusion.

One might argue, however, that this effect could be the result of a general convexity bias and not to prior knowledge about facial geometry. We believe this is less plausible because of results showing the hollow-face illusion is more than a convexity bias for verbal judgments (Hill & Bruce, 1994) coupled with the similarity between the reaching and verbal data (cf. Figure 4 and Figure 6). In fact, general convexity would require stronger assumptions to hold. In particular, it would require that (a) the general convexity bias is stronger in reaching tasks than in the verbal task and (b) the increase in strength was exactly large enough to make the illusion's effect on the two tasks the same.

Also, our findings are consistent with previous research showing that the motor system takes into account prior knowledge about an object in different grasping tasks (e.g., Gordon, 1993; Fikes, Klatzky, & Lederman, 1994; Haffenden & Goodale, 2000).

Not only is prior knowledge used to guide reaches, it can even dominate binocular disparity for the given stimuli, as is shown in Section 3.2. This raises the question, how are binocular disparity and prior knowledge combined?

4.2 Cue combination

The second question we addressed was how prior knowledge and binocular disparity interact when in conflict. A simple strategy would be a winner-take-all approach



Figure 6. Each data point is the average depth estimate for one task, plotted against the average estimate for another task, for one orientation (convex vs. concave) for one participant. If the effect of the illusion on both tasks is the same, the concave data point should lie on the line between the origin and the convex data point. This is true for three of the five participants when comparing the verbal task to either reaching task. The illusion had the same effect on the two reaching tasks for a different group of three of the five participants. Error bars indicate ±1 *SEM*.

in which the visual system relies solely on either binocular disparity or prior knowledge for its depth estimate. Because the hollow-face illusion exists, it is clear that the visual system does not rely solely on binocular disparity. The data in Section 3.2 show that concave faces are estimated to be flatter than convex faces, so the visual system is not relying solely on prior knowledge either. Clearly, depth information from prior knowledge and binocular disparity is being combined in some way.

Integration of prior knowledge with current data has a simple interpretation in terms of Bayesian models of perception. Previous work on surface depth perception has provided strong evidence for a model of depth cue integration that combines information in a statistically optimal fashion (for reviews see, Ernst & Bülthoff, 2004; Bülthoff & Yuille, 1996; Yuille & Bülthoff, 1996; Landy, Maloney, Johnston, & Young, 1995) using Bayesian inference. In these models, cue information is modeled using a likelihood function (the conditional probability of the cue value given a depth) for each cue. Cues and prior information (in the form of probability distributions) are then integrated by multiplying the distributions. In the simplest form of these models, likelihood functions and priors can be modeled as

Gaussian distributions on depth or shape variables. In this case, the optimal estimate (maximum a posteriori) has a particularly simple form-a linear combination of the maximum likelihood depth/shape estimates from each distribution, weighted by its inverse variance (reliability). Linear cue integration models can also serve as useful approximations to optimal statistical inference even when the distributions are not Gaussian (Yuille & Bülthoff, 1996).

A linear cue integration model for our experiment is shown in Equation 1:

$$d = w_p \, d_p + w_b \, d_b + W_{oc} D_{oc} \,, \tag{1}$$

where *d* is the combined depth estimate, d_b is the individual depth estimate from binocular disparity, and d_p represents the depth expected from prior knowledge. w_p and w_b are weights on those individual depth estimates that represent the relative reliabilities. Finally, $W_{oc}D_{oc}$ represents some unknown linear combination of other cues (e.g., pictorial cues, shape from shading, etc.) or priors (e.g., a bias toward a surface smoothness) that may affect perception of our face stimuli.

Note that d_b changes sign between the convex and concave cases. If w_b is small compared to w_p , d will be smaller for concave faces than for convex faces, but will not change sign, so the concave faces will appear to be flatter than the convex faces, but will not be perceived to be concave. This is consistent with our results. However, this is not the only model consistent with these results.

An alternative explanation for our results can be formulated using robust approaches to statistical cue combination (Clark & Yuille, 1990; Maloney & Landy, 1989; Landy, Maloney, & Young, 1991; Shunck, 1989; Sinha & Shunck, 1992). In robust cue combination, data are disregarded if it falls too far outside of expected parameters or if it is inconsistent with other data assumed to be reliable. It is possible that when viewing a convex face, where prior knowledge and binocular disparity are in agreement, w_b has its typical value. However, the conflict between prior knowledge and binocular disparity generated by viewing a concave face may result in the binocular disparity information being "thrown out" as unreliable. In this case, w_b would be set equal to zero. If we also assume that w_n and d_n are the same in the concave case as in the convex case, and that $W_{oc}D_{oc}$ includes a strong bias toward a smooth surface, the new estimate d will be smaller in the concave case, and the face will appear to be flatter than in the convex case. Therefore, this robust statistical approach could also be consistent with our results. Because our data do not test these assumptions, the nature of depth cue combination in the motor system must be resolved by further study.

4.3 Task dependence

Finally, we show that the magnitude of the hollow-face illusion is similar for all three tasks (cf. Figure 4 and Figure 6). This can parsimoniously be explained if we assume that in all tasks the depth estimates are generated by the same mechanism. Our results are consistent with studies that found similar effects of visual illusions on perception, grasping, pointing, and saccades (e.g., Pavan, Boscagli, Benvenuti, Rabuffetti, & Farne, 1999; van Donkelaar, 1999; Franz, Gegenfurtner, Bülthoff, & Fahle, 2000; Dassonville & Bala, 2004) and might help to resolve the current debate on the question of whether motor behavior and perception rely on fundamentally different processing of visual information (e.g., Bridgeman, Kirch, & Sperling, 1981; Aglioti, DeSouza, & Goodale, 1995; for reviews, see Bruno, 2001; Carey, 2001; Franz, 2001; Smeets & Brenner, 2001; Glover, 2002; Goodale & Westwood, 2004).

Kroliczak, Heard, Goodale, and Gregory (in press) have recently described an experiment in which participants were required to "flick" a small target object (a little magnet) off of a location on masks of convex or concave faces. These flicking movements were directed at the real, rather than the illusory, locations of the targets and therefore did not show an effect of the hollow-face illusion. Kroliczak et al. (in press) interpreted their results as consistent with the hypothesis of distinct visual pathways for perceptual judgments versus goal-directed movements. We see, however, two limitations of this conclusion.

First, Kroliczak et al. (in press) did not use ambiguous feedback (as we did in the present study). That is, participants were required to really flick the little magnets from the masks and the magnets were always located at the real, not at the illusory, location on the faces. In consequence, a participant whose motor system was deceived by the hollowface illusion could not perform the flicking at all and should have stopped in mid air, trying to flick unsuccessfully. It seems plausible that such a participant immediately changed the motor strategy to accomplish the task. This could happen in two ways: (a) The participant could try to use any available cue to detect whether the current stimulus is the normal or the hollow face and, in the case of the hollow face, simply move further than the visual input would normally tell the motor system. There were ample cues in this experiment that allowed participants to discriminate between hollow and normal faces. For example, the magnets were always convex such that for the hollow face there was a conflict between the concave shape of the face and the convex shape of the magnets. Also, the faces were illuminated by a little spotlight that was either placed above the normal face or below the hollow face. Such a spotlight creates a brightness gradient, such that its position is detectable by the participant and therefore a discrimination is possible between normal and hollow faces. (b) The participant could weight the binocular information more in this task to detect the real positions of the magnets. (For practical reasons the binocular information was artificially degraded in this study, but this need not interfere with the possibility to utilize it by weighting it more; see our discussion of Bayesian models above.) In summary, a "fair" experimental procedure would require that the target object is either presented at both, the illusory as well as the real positions on the face, or (even better) that flicking is always successful, no matter at which distance the participant attempts to perform the flicking. This is what we achieved by the use of a virtual environment and ambiguous feedback (cf. Figure 2).

Second, the fact that Kroliczak et al. (in press) found no effect of the hollow-face illusion in their flicking task, but did find an effect in a pointing task (which was similar to the flicking task, but required no flicking and was performed slower than the flicking) is interpreted by them as an indication that flicking was controlled by a system other than the slow pointing movements (dorsal vs. ventral streams, respectively). However, an analysis of the computational requirements for various tasks provides another level of explanation for the various ways in which cues may be combined (or rejected) other than this interpretation of two distinct systems. Schrater and Kersten (2000) used decision theory to show that cue combination for optimal depth estimation depends crucially on the representation of depth (see Geisler & Kersten, 2002, for a simple illustration of decision theory for perceptual estimations). In particular, the best estimate of depth of a target depends on

how (not just whether) information about a background surface is represented. Reaching movements could depend on whether the target object is treated as distinct from the surface or as part of the surface. This, in turn, could depend on visual factors (whether a target is in contact, not in contact, or a surface marking) and also on task prerequisites (e.g., "flicking" implies removability, touching does not). In addition to decision theoretic constraints, dynamical constraints with respect to the goal of the reach should also play an important role in determining visual motor trajectories. The kinematics, up to the point of expected contact, can depend on the expected consequences beyond the time of contact. For example, if a target is being touched with a movement perpendicular to a background surface, any follow through of the movement would be blocked by the surface, and thus background surface depth is an important piece of information. If it is being flicked, it is free to move tangential to the surface, and the background surface depth is less crucial. Task constraints may modulate cue integration through changes in attentional allocation.

Conclusion

Using hollow faces as a target for distance estimations, we have shown that prior knowledge of object shape can dominate shape from binocular disparity information in reaching tasks, as well as in verbal tasks. The shape estimates from the two sources of information are combined, rather than one being thrown out as completely unreliable. The resulting shape estimates are similar for both verbal and reaching tasks, which is what we would expect if the same cue combination strategy is being used for the reaching and the verbal tasks.

Acknowledgments

This work was first presented at the 2001 Vision Sciences Society Conference, Sarasota, Florida (cf. Hartung, Franz, Kersten, & Bülthoff, 2001). The work was supported by National Institutes of Health Grants R01 EY11507 and R01 EY015261-01, the Max Planck Society, and grant FA 119/15-2 from the Deutsche Forschungsgemeinschaft.

Commercial relationships: none. Corresponding author: Volker H. Franz. Email: volker.franz@psychol.uni-giessen.de. Address: University of Giessen, Otto-Behaghel-Strasse 10F 35394, Giessen, Germany.

References

Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5 (6), 679- 685. [PubMed]

- Belhumeur, P., Kriegman, D., & Yuille, A. (1999). The basrelief ambiguity. International Journal of Computer Vision, 35(1), 33-44.
- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. SIGGRAPH'99 Conference Proceedings, 187-194.
- Bridgeman, B., Lewis, S., Heit, G., & Nagle, M. (1979). Relation between cognitive and motor-oriented systems of visual position perception, *Journal of Experimental Psychology: Human Perception and Performance*, 5, 692-700. [PubMed]
- Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception & Psychophysics*, 29, 336-342. [PubMed]
- Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception & Psychophysics 59*, 456-469. [PubMed]
- Bruno, N. (2001). When does action resist visual illusions? Trends in Cognitive Sciences, 5(9), 379-382. [PubMed]
- Bülthoff, H. H., & Yuille, A. L. (1996). A Bayesian framework for the integration of visual modules. In J. McClelland & T. Inui (Eds.), Attention & performance XVI: Information integration in perception and communication (pp. 49-70). Cambridge: MIT Press.
- Clark, J. J., & Yuille, A. L. (1990) Data fusion for sensory information processing systems, Boston: Kluwer Academic Publishers.
- Carey, D. P. (2001). Do action systems resist visual illusions? Trends in Cognitive Sciences, 5(3), 109-113. [PubMed]
- Dassonville, P., & Bala, J. K. (2004). Perception, action, and Roelofs effect: A mere illusion of dissociation. *PLoS Biology*, 2(11), 1936-1945. [PubMed][Article]
- Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8(4), 162-169. [PubMed]
- Fikes, T. G., Klatzky, R. L., & Lederman, S. J. (1994). Effects of object texture on precontact movement time in human prehension. *Journal of Motor Behavior*, 26, 325-332. [PubMed]
- Franz, V. H. (2001). Action does not resist visual illusions. *Trends in Cognitive Sciences*, 5(11), 457-459. [PubMed]
- Franz, V. H., Gegenfurtner, K. R., Bülthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science*, 11(1), 20-25. [PubMed]
- Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience*, *5*, 508-510. [PubMed]
- Glover, S. (2002). Visual illusions affect planning but not control. *Trends in Cognitive Sciences*, 6(7), 288-292. [PubMed]

- Goodale, M. A., Jakobson, L. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, 32, 1159-1178. [PubMed]
- Gordon, A. M., Westling, G., Cole, K. J., & Johansson, R. S. (1993). Memory representations underlying motor commands used during manipulation of common and novel objects. *Journal of Neurophysiology*, 69, 1789-1796. [PubMed]
- Goodale, M. A., & Westwood, D. A. (2004). An evolving view of duplex vision: Separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology*, 14, 203-211. [PubMed]
- Gregory, R. L. (1973). The confounded eye. In R. L. Gregory & E. H. Gombrich (Eds.), *Illusion in nature and art* (pp. 49-96). London: Duckworth.
- Haffenden, A. M., & Goodale, M. A. (2000). The effect of learned perceptual associations on visuomotor programming varies with kinematic demands. *Journal of Cognitive Neuroscience*, 12(6), 950-964. [PubMed]
- Hartung, B., Franz, V. H., Kersten, D., & Bülthoff, H. H. (2001). Is the motor system affected by the hollow face illusion? [Abstract] Journal of Vision, 1(3), 256a, http://journalofvision.org/1/3/256/, doi:10 .1167/1.3.256.
- Hespanha, J., Dodds, Z., Hager, G. D., & Morse, A. S. (1999). What tasks can be performed with an uncalibrated stereo vision system? *The International Journal of Computer Vision*, 35(1), 65-85.
- Hill, H., & Bruce, V. (1993). Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. *Perception*, 22, 887-897. [PubMed]
- Hill, H., & Bruce, V. (1994). A comparison between the hollow-face and 'hollow-potato' illusions. *Perception*, 23, 1335-1337. [PubMed]
- Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception and Performance, 22, 986-1004. [PubMed]
- Kroliczak, G., Heard, P., Goodale, M. A., & Gregory R. L. (in press). Dissociation of perception and action unmasked by the hollow-face illusion. *Cognitive Brain Research*.

- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion, *Vision Research*, 35, 389-412. [PubMed]
- Landy, M. S., Maloney, L. T., & Young, M. J. (1991). Psychophysical estimation of the human depth combination rule. *Proceedings of the SPIE*, 1383, 247-254.
- Langer, M. S., & Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, 30, 403-410. [PubMed]
- Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of the SPIE*, 1199, 1154-1163.
- Milner, A. D., & Goodale, M. A. (1995). The visual brain in action. Cambridge: Oxford University Press.
- Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., & Farne, A. (1999). Are perception and action affected differently by the Titchener circles illusion? *Experimental Brain Research*, 127, 95-101. [PubMed]
- Schrater, P. R., & Kersten, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision*, 40(1), 73-91.
- Schunck, B. G. (1989). Robust estimation of image flow. *Proceedings of the APIE*, 1198, 116-127.
- Sinha, S. S., & Schunck, B. G. (1992). A two stage algorithm for discontinuity-preserving surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 36-55.
- Smeets, J. B. J., & Brenner, E. (2001). Action beyond our grasp. *Trends in Cognitive Sciences*, 5(7), 287.
- Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. Vision Research, 36, 1761-1771. [PubMed]
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123-161). Cambridge: Cambridge University Press.
- van Donkelaar, P. (1999). Pointing movements are affected by size-contrast illusions. *Experimental Brain Research*, 125(4), 517-520. [PubMed]