Journal of Vision (2006) 6, 75–86 http://journalofvision.org/6/1/7/

Shape recognition alters sensitivity in stereoscopic depth discrimination

Hongjing Lu
Department of Psychology, University of California,
Los Angeles, CA, USA

Bosco S. Tjan
Department of Psychology and Graduate Program
of Neuroscience, University of Southern California,
Los Angeles, CA, USA

Zili Liu
Department of Psychology, University of California,
Los Angeles, CA, USA

A fundamental question in visual perception is to characterize how information from sensory input is integrated with prior probabilities. The role of prior probabilities is controversial for elementary visual processes, which are often believed to be immune from higher-level influences. In this paper, we demonstrate such influences. We tested human observers' abilities to discriminate stereoscopic depth defined by points embedded in a biological pattern—a human figure. Our results indicate that the internal representation of a walking human figure imposed constraints on depth discrimination of a static stimulus. This constraint was manifested when the stimulus was recognized as a human figure. When the expected human figure configuration (forearms having equal length) was inconsistent with the sensory input information, discrimination of forearm lengths was impaired. In contrast, when there was no inconsistency (the hand–hip distance on the left was not expected to be equal to that on the right), discrimination between the two distances was improved, presumably because the human figure configuration provided a more accurate frame of reference for stereoscopic depth. Both the impairment and improvement were due to changes of discrimination sensitivity rather than decision bias. Our findings support the view that visual perception is an inference process constrained by Bayesian prior probabilities.

Keywords: Bayes, bias, biological motion, binocular disparity, computation, depth, discrimination, stereo, sensitivity

Introduction

Ever since Helmholtz (1866/1924) characterized perception as a process of "unconscious inference," an important problem in visual perception has been to validate or refute this inference process. Inference is necessary, according to Helmholtz, because stimulus information for perception is often ambiguous or underconstrained. Prior preferences or expectations are therefore needed to further constrain the possible interpretations to achieve a unique and unambiguous percept. To this end, significant progress has been made that supports the inference theory in perception (Feldman, 2000; Hartung, Schrater, Bülthoff, Kersten, & Franz, 2005; Kersten, Bülthoff, Schwartz, & Kurtz, 1991; Kersten, Knill, Mamassian, & Bülthoff, 1996; Mamassian & Landy, 1998, 2001; Nakayama & Shimojo, 1992; Weiss, Simoncelli, & Adelson, 2002).

The priors that have been so far studied are general purpose, in the sense that they characterize what the perceptual system believes is likely before stimulus information is considered. A different type of priors, which have not been considered in visual perception, is more special purpose, partially stimulus dependent, and sometimes referred to as model selection (Grenander, 1993). As an example, a system may have a number of priors or models. Depending on the context and preliminary processing of stimulus information, the system may switch from one prior to another to achieve a more reliable and accurate inference. To use specific rather than generic priors, this hypothesis testing and retesting process necessarily requires interaction with and preliminary processing of stimulus information. A concrete (although speculative) example is recognition of moony (two-tone) faces. When viewing an image of a specific moony face for the first time, a generic prior is unlikely to be helpful. If an observer instead considers specific face models of various sizes and orientations, particularly if aided with hints and instructions, recognition is more likely to succeed. In comparison, if the face is replaced by a two-tone generic object, the required specific priors are prohibitively too many for recognition to be possible (Moore & Cavanagh, 1998).

In this paper, we provide empirical support to the plausibility of the model selection in perception. We studied stereoscopic depth discrimination of dots and found that observers' discrimination thresholds were changed after the dots were recognized as forming a human figure. When expectations based on human geometry (e.g., forearms are equally long) contradicted stimulus disparity information,

doi: 10.1167/6.1.7 Received December 17, 2004; published January 27, 2006 ISSN 1534-7362

discrimination became worse. In contrast, when this contradiction was removed, discrimination was improved after recognition, possibly because recognition of the otherwise meaningless dots as a human figure improved the representation of the dots' positions. We developed a simple Bayesian model to account for both the decrease and increase in discrimination performance.

Our approach is similar to that of analyses of cue interactions in the literature (Clark & Yuille, 1990; Landy, Maloney, Johnston, & Young, 1995), although shape recognition is a unique cue in our study, as will be clear. Because shape recognition is often considered as a "higher-level" visual function, its interaction with "lower-level" cues, such as shading and binocular disparity that are often thought to subserve shape recognition, is sometimes referred to as a "top-down" interaction (Bülthoff, Bülthoff, & Sinha, 1998; Moore & Cavanagh, 1998). However, because there is no unequivocal evidence that shape recognition takes place after stereoscopic depth discrimination, we will not in this paper use this term.

Previous work

Configural effects impairing stereo discrimination

Depth discrimination had previously been assumed to be limited by disparity detectors in the primary visual cortex V1 (Farell, 1998; Morgan & Castet, 1997). More recent neurophysiological evidence, however, suggests that V2, V3, and MT all play an important role in processing disparity information (Adams & Zeki, 2001; Backus, Fleet, Parker, & Heeger, 2001; DeAngelis, Cumming, & Newsome, 1998; DeAngelis & Newsome, 1999, 2004; Grunewald, Bradley, & Andersen, 2002). Therefore, it is appropriate to consider configural effects in stereoscopic discrimination as being due to cue integration and competition (Clark & Yuille, 1990; Landy et al., 1995). As described below, this conception can explain a number of findings.

A classic example of stereoacuity discrimination is the relative depth judgment of two parallel vertical bars (McKee, 1983; Mitchison & Westheimer, 1984). McKee (1983) and Mitchison and Westheimer (1984) found that when the vertical bars were connected by two horizontal bars to form a square, discrimination threshold was greatly elevated. This result can be understood as being due to cue competition between shape-from-stereo and shape-from-perspective. The monocular linear perspective cue from the stimulus indicated a square in the fronto-parallel plane, thereby impairing depth discrimination of the two vertical bars. The remaining findings from the study of Mitchison and Westheimer (1984) can be similarly explained. Threshold of depth discrimination between two columns of dots was greatly elevated when additional flanking columns were added because the entire stimulus became a square of regular grid in the fronto-parallel plane according to the monocular linear perspective cue, although the flanking columns provided additional stereoscopic depth cues for the dot plane that was slanted away from the fronto-parallel plane. Even when the two vertical bars in the study by McKee (1983) were connected by a horizontal bar to form a letter H, discrimination threshold was similarly elevated. Likewise, when discriminating stereoacuity between two dots, threshold was elevated when a third dot was added that was collinear with the first two (Fahle & Westheimer, 1988).

In another manipulation (McKee, 1983), three parallel vertical bars were presented in stereo and human participants judged whether the central bar was in front of or behind the plane defined by the other two bars. Discrimination threshold was markedly elevated when two additional vertical bars were added to the plane, one above the central bar and one below, such that these three bars could be misperceived as being collinear. The collinear cue therefore competed with the stereo disparity cue to make the depth judgment of the central bar more difficult. This result was subsequently extended to stereoscopic discrimination of surfaces (Hou, Lu, Zhou, & Liu, in press; Liu, Jacobs, & Basri, 1999; Yin, Kellman, & Shipley, 2000). That is, when two surfaces could by geometric arrangement perceptually connect with one another to form a single surface behind occlusion (amodal completion), it became more difficult to stereoscopically discriminate them in depth.

The plausibility of the cue interaction account can be further illustrated by a recent study of surface stereoacuity discrimination (Vreven, McKee, & Verghese, 2002). The stereoscopically defined surface was either planar or curved and was depicted either with constant luminance (stereo disparity information provided at the boundary) or with random dots. A small bar serving as the stereo depth probe was positioned either in front of or behind the surface (or, when curved, the surface peak). It was found that discrimination was better for a planar than for a curved luminance surface, better for a smaller than for a larger portion of a curved luminance surface, and better for a curved surface defined by random dots than by constant luminance.

Although Vreven et al. (2002) offered separate accounts for these effects observed for the planar luminance surface, the curved luminance surface, and the random-dot surface, it appears that a single account based on cue competition between shape-from-stereo and shape-from-shading may explain all the findings. Specifically, constant luminance, as a shape-from-shading cue, generally indicates a planar surface. Therefore, when the left and right boundaries of a surface are defined by binocular disparity to indicate curvature in space, and at the same time the entire surface is shaded to indicate a planar surface with unspecified orientation in depth, the surface thereby created inconsistent cues from stereo and shading. It is impossible to quantify what the resultant surface might perceptually be without additional assumptions. Nevertheless, uncertainties associated with estimating the absolute depth, local surface orientation, and local curvature at any point are greater for a curved surface with constant luminance than a planar or random-dot surface. This is primarily because the depth location and orientation of any local planar component implied by the shading in the constant-luminance case are largely underconstrained.

As a result, this account can explain why the planar surface (no cue conflict) yielded a lower threshold than the curved luminance surface (cue conflict). It can also explain why a smaller portion of the curved luminance surface yielded a lower threshold than a larger portion of the same surface. This result follows because a smaller portion, albeit with a smaller amount of disparity information, would also yield less cue conflicts, and at the same time, depth interpolation at the surface peak would be less error prone because the peak was closer to the boundaries. Finally, this account can also explain why the random-dot surface yielded a lower threshold. Because this surface was now defined by random dots, the background of constant luminance was no longer perceived to belong to the surface, and hence generated no cue conflict. Moreover, additional disparity information became available at every dot location.

Richer stimulus information facilitating stereo discrimination

In addition to the impairment identified in the studies above, improvement in depth discrimination has also been found. However, the evidence appears to suggest that this improvement was mainly due to additional, rather than better organized and structured, stimulus information. For instance, discrimination threshold of a test line was lowest when the reference line was in the fixation plane (Badcock & Schor, 1985; Blakemore, 1970; McKee, Levi, & Bowne, 1990; Ogle, 1953). Glennerster and McKee (1999) showed that when a disparity-defined plane (fronto-parallel or slanted) was added to a stereoacuity stimulus, depth discrimination appeared to be made relative to this plane as a reference. The addition of such a plane improved discrimination sensitivity. Kumar and Glaser (1992) further showed that adding additional reference lines to a single test line stimulus lowered discrimination threshold by a factor of 10. In fact, four reference lines were better than two, which in turn were better than one. Moreover, the improvement was greater when the test line was between two reference lines in depth (see also Westheimer & Levi, 1987). These findings suggest that explicit, multiple, and redundant references in depth improved precision in encoding relative depths. In other words, richer depth information (even if redundant) appeared sufficient to improve subtle depth discrimination, whereas stimulus configuration was not critical.

Indeed, a study by Wallace and Mamassian (2004) found that any difference in depth discrimination between transparent and opaque surfaces was entirely due to inherent correspondence uncertainty in the stimulus rather than limits on visual processing. Specifically, two random-dot stimuli were presented in stereo, one forming a plane in front of the fixation point, the other behind. Human participants discriminated which plane was closer to the fixation. When the two stimuli were presented simultaneously, stereo transparency was perceived. In contrast, when the two stimuli were presented sequentially, one opaque plane was perceived at a time. Human discrimination threshold was found to be higher for the transparency than for the opaque condition. These threshold differences were comparable for an ideal observer, showing that statistical efficiency was equal between the two conditions. This equality implied that the human threshold difference was completely accounted for by the difference in the input stimuli, and not by the limits of visual processing.

Shape recognition and bias versus sensitivity

Relative to the geometric (binocular disparity, linear perspective, colinearity, and coplanarity) and photometric (shading) cues discussed so far, the cue of shape recognition is unique. It is considered as a "cue" because it imposes structure and meaning on the input stimulus after recognition. The imposition may be so strong that "abnormalities" of other cues, such as stereo, are perceptually ignored, as demonstrated by Bülthoff et al. (1998). These authors used animated biological motion stimuli (i.e., a movie of a human figure formed by representing only the positions of the joints by bright dots; Johanssen, 1973) to demonstrate that recognition of a familiar three-dimensional (3D) human figure could override objective stereoscopic depth information. Participants viewed a 3D point-light walker that was scrambled in depth and were asked to rate the display in terms of goodness as a human figure. It turned out that the depth-distorted walker was rated as good a human figure as the intact walker. Thus, 3D stereoscopic depth perception, according to these authors, was completely overridden by the expected 3D shape of a human figure that was recognized from the 2D projections of the dots.

However, because a subjective rating was used as the dependent measure by Bülthoff et al. (1998), it is unclear whether the shape recognition altered response bias or discrimination sensitivity of stereoscopic depth. A good rating may have been either due to the availability of a human figure interpretation for the depth-randomized points (a response bias) or due to depth randomization being more difficult to detect (a sensitivity change).

In fact, "higher-level" cues, historically termed "mental set," have long been observed. These studies can be almost invariably characterized as altering biases but not discrimination sensitivity. For instance, the perceived depth ordering of the ambiguous Necker cube was found to be influenced by both instructed fixations and observers' intentions (Hochberg & Peterson, 1987; Peterson & Hochberg, 1983). Peer pressure from a group also biased towards group norms an observer's perceived autokinetic effect (Sherif, 1935) and perceived length of a line (Asch, 1955) (see also Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Fisher, 1968; Gregory, 1978; Mussap & Levi, 1995; Remez, Rubin, Pisoni, & Carrell, 1981; Suzuki & Cavanagh, 1998).

Although it is unclear whether shape recognition (of a human figure) altered stereoscopic discrimination sensitivity in the study of Bülthoff et al. (1998), there is indirect evidence that it may be possible. This is because Tadin, Lappin, Blake, and Grossman (2002) have demonstrated that an upright point-light human walker (more familiar) gave rise to better temporal discrimination than an inverted point-light walker (less familiar). More specifically, in a 2D display of an animated human walker, in which the point lights were replaced by Gabors that were changing phases, participants judged more accurately for an upright than for an inverted walker whether the phase change was synchronous. It is likely that this difference in accuracy could translate into a difference in discrimination sensitivity because systematic bias was not expected in this forced-choice task. Interestingly, when the human figures were static, the overall performance was improved but no difference was found between the upright and inverted recognizable figures by the experienced participants. Tadin et al. (2002) concluded that the upright animated walker served as a more efficient reference frame for the Gabors so that timing discrimination was easier. This evidence is indirect to infer depth discrimination with and without shape recognition, however, because the participants were experienced and informed and could recognize the human figure even when it was inverted (Tadin, 2004, personal communication). The evidence is also indirect because the task was temporal discrimination.

In the current study, we demonstrate a difference in depth discrimination sensitivity (as opposed to response bias) between upright and inverted static point-light humans using a stimulus similar to that used by Bülthoff et al. (1998). We show that the observed difference depends on whether the stimulus is recognized as a human figure. We also show that sensitivity can either decrease or increase, depending on whether the recognized shape is inconsistent or consistent with stereoscopic information.

Experiment 1: Forearm length discrimination with recognition

Method

Participants were required to discriminate 3D distances between two pairs of stereoscopically presented stationary points. As stimuli, we used a 3D human figure consisting of 13 points (Figure 1). These stimuli were the same as those used by Bülthoff et al. (1998), except that we added a head point and adjusted the 3D positions of the points to ensure perfectly symmetric lengths of corresponding limbs (by setting each forearm length to the mean length of the original left and right forearms). Hand and elbow joints were selected for stereo distance discrimination in Experiment 1. The two point lights representing one forearm were marked red, and those representing the other forearm were marked green. The remaining point lights were blue. Participants judged in a static stereoscopic display whether the 3D stereo distance between the red pair was larger than between the green pair. Each stimulus was presented such that the 2D projected distances were equal between the red pair and between the green pair. Therefore, participants could not base their discrimination on 2D distances. Further, judging the difference between the two 3D distances became equivalent to judging the relative depth difference. However, given that the 3D distance was meaningful after recognition (forearm length in Experiment 1) and was central to the hypothesis being tested, we will use the term "3D distance" in this paper. "3D distance" was also the term used in experimental instructions.

The experiment had two sessions, each of which consisted of a psychometric function measurement, using the method of constant stimuli. In the first session, the stimulus was static and inverted, so that participants who had never seen biological motion stimuli could not recognize the stimulus as a human figure. This nonrecognition was confirmed by each participant at the end of the experiment. After the first session, an animation of the intact upright 3D human walker was presented in stereo for 60 s so that it was readily recognized. After recognition, the psychometric function was again measured with the static stimulus turned upright so that the human figure could be recognized on every trial.

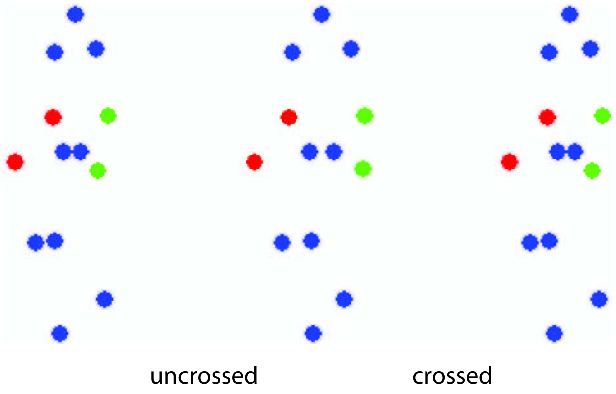

Figure 1. A stereogram of an upright human figure with a red forearm and a green forearm. This stereogram contains two stereo pairs, one at crossed disparity and the other at uncrossed disparity. When the stereogram is fused, the green forearm is longer than the red forearm.

On each trial, a stimulus was presented stereoscopically with orthogonal projection against a gray background (22.6 cd/m^2) on a computer display (Silicon Graphics, Inc.) that was equipped with stereo shutter goggles (60 Hz). The viewing distance was 57 cm. The height and width of the stimulus were 13 and 6 deg in visual angle, respectively. The luminance of the red, green, and blue points was 19.9 cd/m^2. Each point subtended 0.03 deg in visual angle. Each stimulus was shown for 4 s. Participants discriminated between the red and green 3D distances by pressing one of two response keys. No feedback was provided. There was a 1-s pause between a response and the next trial. The simulated 3D length difference between the forearms at the given viewing distance was selected from one of the following six levels: 0.34, 0.67, 1.33, 3.33, 5.34, and 7.34 mm. The color and length assignments to the forearms were randomized. The mean of the two lengths was held constant at 16 mm.

Each session of psychometric measurement consisted of 216 trials that were randomly interleaved, with 36 trials per level. These 36 trials were created from 36 poses of the human walker that were obtained as follows. The original 3D point-light human walker consisted of a sequence of 20 poses. The first 17 were used to generate 16 additional poses by interpolating between adjacent poses, using angular bisection. The resulting 36 poses in 3D were used for every level of the 3D length difference. For each pose, the 3D distance between one pair of dots (e.g., red) was lengthened, and that between the other pair (e.g., green) was shortened to reach the desired level of the 3D length difference. The animated camera then rotated around the vertical axis to find a viewpoint that equalized the projected 2D lengths. When no such viewpoint could be found, one forearm would be rotated around its elbow in the plane of the full arm to increase the search space. If unsuccessful, both forearms would be rotated. If still unsuccessful, one or both full arms would be rotated around the vertical axis at the corresponding shoulder joints to further increase the search space. In all cases, this procedure proved sufficient to find the desired 3D length difference while keeping the projected 2D lengths equal. We emphasize here that an identical set of stimuli was used for both sessions in this experiment.

For the animation movie, the original 20-pose sequence of the human walker was shown, each pose for 40 ms. From one pose to the next, the viewpoint was rotated around the vertical axis by 1 deg. The entire movie was played for 60 s. Participants were not required to perform any task when viewing the animation. There was a 30-s break between the first session and the animation and between the animation and the second session. A computer beep prompted the start of the animation and each session. The entire experiment lasted for about 40 min.

Seventeen undergraduate students from University of California Los Angeles (UCLA) participated in Experiment 1. When debriefed at the end of the experiment, all participants reported that they had recognized the human figure during the animation and the second session. All participants in this and subsequent experiments reported that they had not previously seen any biological motion stimuli.

Results and discussion

We used a cumulative Gaussian to fit the psychometric curve, in which the horizontal axis was the simulated 3D length difference between the red and green dot pairs, and the vertical axis was the proportion of responses indicating that the length between the red points was larger. With respect to signal detection theory (Green & Swets, 1974), the mean of the Gaussian represents the bias (if different from zero) whereas the standard deviation represents the reciprocal of sensitivity (because it corresponds to the amount of decision noise). It should be noted that sensitivity thus defined is not d'. We also analyzed the data by computing discrimination sensitivity d' and bias β at each of the six length differences. A standard ANOVA yielded results completely consistent with the cumulative Gaussian analysis. This was true for Experiments 2 and 3 also.

In Figure 2, the psychometric curve was calculated by averaging the fitting means and standard deviations across participants. The data in Figure 2 indicate that the slope of the psychometric function became shallower from the first to the second session. This result indicates that sensitivity (1/*SD*) decreased after recognition (from an *SD* of 3.36 to 4.28 mm, $t(16) = 2.23$, $p = .041$, all t tests reported in this paper were two tailed). The change in response bias was nonsignificant, $t(16) = 0.79$, $p = .44$; mean bias = −0.29 mm. The nonsignificant negative value of the mean bias indicates that participants slightly preferred judging the distance between the red dots as longer. This trend might have resulted from the fact that "red" corresponded to the right response button in the experiment, and the majority of the participants (14/17) were right-handed.

Our results imply that, after recognition, participants expected the forearms to be approximately equal in length, an expectation that impaired discrimination of 3D lengths by reducing sensitivity. Note that if any participants recognized the inverted human figure in the first session (contrary to their report upon debriefing at the end of the experiment), the observed change in sensitivity would likely be less. Therefore, the reduced sensitivity observed here was based on a conservative measure.

Experiment 2 (control): Forearm length discrimination without recognition

One possible interpretation of the results in Experiment 1 was that the sensitivity reduction was due to session

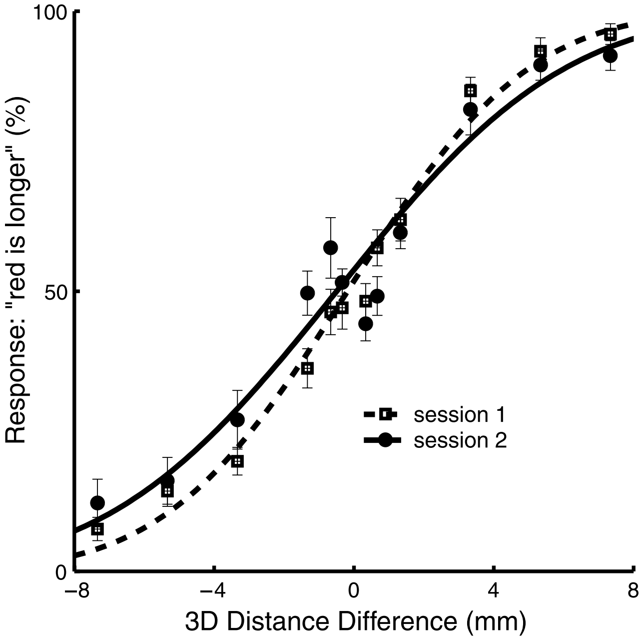

Figure 2. Proportion of responses that the red limb was longer than the green limb as a function of the difference between their 3D lengths. Data are means ± *SEM*. Square symbols show results for Session 1; circle symbols show results for Session 2. The psychometric function for each participant was individually fitted with a cumulative Gaussian, where its standard deviation represents the reciprocal of discrimination sensitivity and its mean represents response bias. After recognition of the dots as a human figure, the slope of the psychometric curve was shallower, indicating sensitivity was reduced whereas bias changed little. The psychometric curves were plotted using the average of the parameters (standard deviations and means of a cumulative Gaussian) obtained from individual participants.

repetition rather than to the intervening recognition of the human figure. Another possible interpretation was that the results were completely due to figural inversion from Session 1 (inverted) to Session 2 (upright). Experiment 2 was designed to control for these possibilities by repeating the two sessions in Experiment 1 without the intervening walker movie that would otherwise trigger recognition of the human figure.

Seventeen UCLA undergraduates (13 right-handed) participated, who had not participated in Experiment 1.

Results and discussion

We found no significant change in observers' sensitivity (*SD,* 1/sensitivity; changed from 3.92 to 3.86 mm, $t(16) = 0.13$, $p = .90$). The change of bias was also nonsignificant (bias changed from −0.26 to −0.32, $t(16) = 0.29$, $p = .78$; see Figure 3).

The fact that little effect was found when no intervening movie was shown indicates that the effects observed in Experiment 1 were more likely due to shape recognition

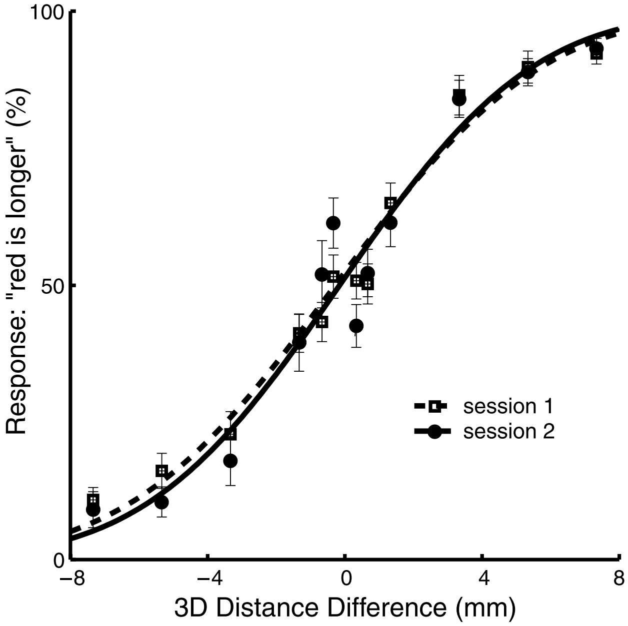

Figure 3. Proportion of responses that the red limb was longer than the green limb as a function of the difference between their 3D lengths. Without recognition of the human figure (when the animation movie was not shown), discrimination sensitivity and bias changed little from Session 1 to Session 2.

triggered by the movie rather than session repetition or to figural inversion. In Experiment 3 (see below), which also used figural inversion and the intervening movie, effects opposite to Experiment 1 were found, providing additional evidence that figural inversion per se was unlikely to account for all the effects.

Experiment 3: Hand–hip distance discrimination with recognition

If the discrimination in Experiment 1 was impaired because participants expected, after recognition, that the forearms of a human figure would be equal in length, one would expect the effect to disappear when the two 3D distances to be discriminated were no longer expected to be equal. Further, we hypothesized that depth discrimination after recognition may even improve when the equal length expectation was irrelevant. An upright human figure, as opposed to an inverted and unrecognized one, might be represented by the visual system more accurately and efficiently and therefore would provide a more precise frame of reference to represent the 3D positions of multiple components (Tadin et al., 2002). We tested this hypothesis in Experiment 3 using the discrimination of hand–hip 3D distances. Because a hand and a hip are attached to separate limbs that move quasi-independently, the left and right hand–hip distances should not be expected to be equal.

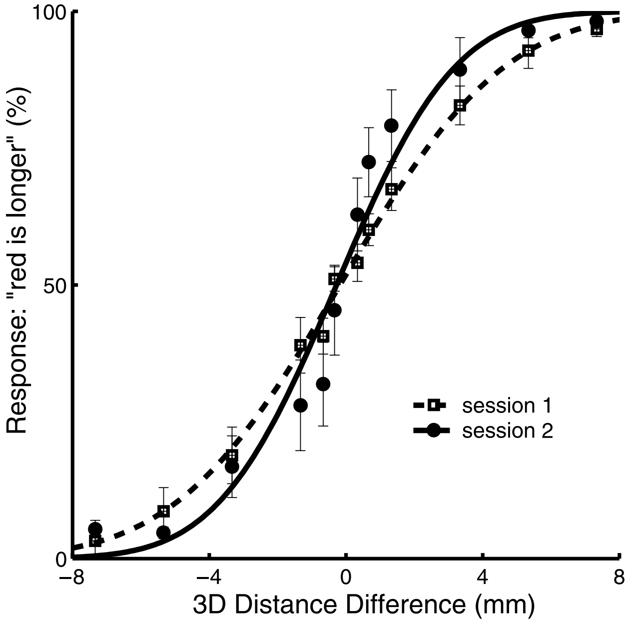

Figure 4. Proportion of responses that the distance between the red dots was larger than that between the green dots as a function of the difference between the 3D distances. Hand–hip distances were not expected to be equal for a human walker. After recognition, discrimination sensitivity increased whereas bias changed little.

Method

The stimuli were generated similarly as in Experiment 1, except that hand–elbow pairs were replaced by hand–hip ones. Specifically, for each human pose in 3D, one hand point was rotated around its elbow point in the plane defined by the full arm to lengthen its distance to the corresponding hip point. The other hand point was likewise rotated to shorten its distance to the hip point. The animated camera then rotated around the vertical axis to find a viewpoint that equalized the projected 2D distances. When no such viewpoint could be found, one hand point would be rotated around its corresponding elbow–hip axis to increase the search space. If unsuccessful, both hand points would be so rotated. If still unsuccessful, one or both full arms would be rotated around the vertical axis at the corresponding shoulder joints to further increase the search space. The hand–hip distances would be re-adjusted accordingly. In all cases, this procedure proved sufficient to find the desired 3D distance difference while keeping the projected 2D distances equal. To ensure a fair comparison of discrimination with regard to the Weber–Fechner's law, the mean 3D hand–hip distance of each stimulus was held the same as in Experiments 1 and 2. Seventeen UCLA undergraduates (12 right-handed) who did not take part in Experiments 1 or 2 participated.

Results and discussion

Discrimination sensitivity of hand–hip distance improved after recognition, without changes in bias. Specifically, standard deviation (1/sensitivity) of the psychometric functions was reduced from 3.04 to 1.91 mm, $t(16) = 2.67$, $p = .02$; whereas the bias did not change, $t(16) = 0.06$,

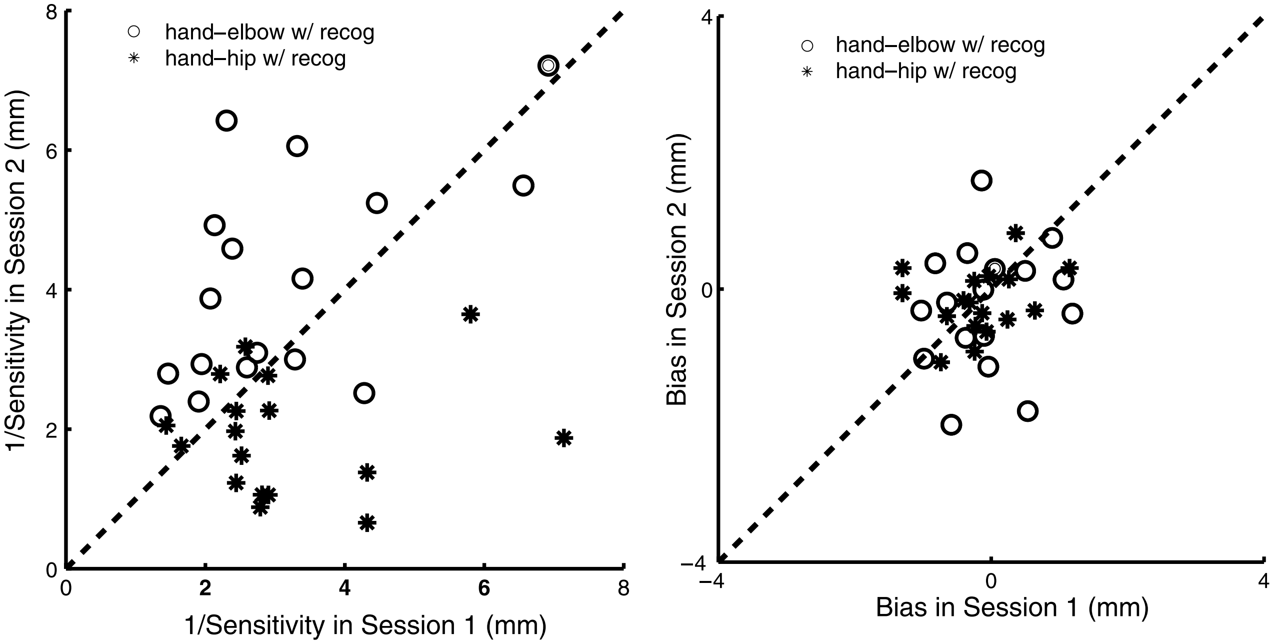

Figure 5. Scatter plot of data for individual participants in Experiment 1 (forearm length discrimination with recognition) and Experiment 3 (hand–hip distance discrimination with recognition). Left: fitted standard deviation (1/sensitivity); right: fitted mean (bias).

p = .95; mean bias = −0.01 mm (see Figure 4). We further verified the hypothesis that the standard deviations (1/sensitivity) from Session 1 of all the three experiments were approximately equal, $F(2,48) = 0.94 < 1$. Similarly, we found that the biases in Session 1 of these experiments were approximately equal (mean = −0.22 mm), $F(2,48) = 0.04 < 1$. No reliable difference in bias was found in Session 2 of these experiments either ($F < 1$). Figure 5 shows individual participant's sensitivity and bias in Experiments 1 and 3.

General discussion

Our results demonstrate that a basic perceptual function—stereoscopic depth discrimination—was influenced by prior knowledge of an ecologically significant shape, namely, the human form. Further, our findings indicate that such influence could either impair or facilitate stereoscopic discrimination sensitivity, without affecting response bias. The change in sensitivity was likely due to different expectations about the geometry of a human figure. Using closely matched discrimination tasks and stimuli, we were able to produce either impairment or facilitation of stereoscopic 3D distance discrimination after participants recognized the human figure. The key difference responsible for the direction of the influence involved the ecological aspects of body geometry: forearms (hand–elbow) are presumed equal in length, whereas hand–hip distances are not. For our stimuli, left and right hand–elbow distance discrimination was impaired by the expectation of equal length, whereas hand–hip distance discrimination was aided by the knowledge of the human figure that served as a reference frame. Before recognition, distance discrimination between the red points and green points may be reasonably considered as a bottom-up process, with the remaining blue points irrelevant. After recognition, all points became components of a meaningful and familiar human figure that influenced 3D distance discrimination between a subset of the points.

From a Bayesian point of view, the prior probabilities of the depth positions of the points were uninformative before recognition (e.g., each point's depth position corresponded to a Gaussian with a large variance). Estimation of the 3D positions of the red and green points was primarily determined by bottom-up sensory processes expressed by the likelihood function for calculating binocular disparity. After recognition, in contrast, model selection would choose prior probability distributions of the points' depths that were no longer uninformative. Instead, the distribution of each point's depth was localized according to the expected position of the point's corresponding body part (e.g., a Gaussian with a finite variance). We note that the prior probability used here is not in the usual sense of the term, rather it is a probabilistic function conditioned on recognizing the stimulus as a human. Specifically, recognition here is defined as being able to label each point relative to a human body (e.g., left hand, left elbow) and consequently knowing the point's 2D (but not necessarily depth) position.

After recognition, the blue points also became informative because they provided additional reference to further reduce uncertainties about the prior expected 3D positions of the points. Because the inference regarding the 3D distance is determined by the product of the prior and the likelihood, the qualitative role of the prior is twofold. For clarity of illustration, we assume that all distributions are Gaussians. It is straightforward to show that (1) if the peaks of the prior and the likelihood are separated, the peak of the posterior distribution is located between them; (2) the variance of the posterior is smaller than that of the likelihood, so long as the variance of the prior is finite. The consequence of the localized prior predicts increased sensitivity in the hand–hip distance discrimination and decreased sensitivity in hand–elbow discrimination.

In what follows, we provide a qualitative account. The more detailed proof can be found in the Appendix, which was originally developed in Hou et al. (in press). Given that the 2D projected distances were identical between the red points and between the green points, 3D distance discrimination was formally equivalent to relative depth discrimination. We therefore consider only the depth dimension.

In the hand–hip distance discrimination, because there was no inconsistency between the *a priori* expected location of a point and its sensory information, the peaks of the prior and the likelihood coincided. Therefore, the posterior estimation of each point's location was more accurate, after recognition—the variance of the distribution becomes smaller after recognition. Because distance estimation between two points has a probability distribution that is the convolution of the distributions of the two points' positions, distance estimation was therefore more accurate after recognition. According to signal detection theory, when the interval between two Gaussians (here representing the two distance estimations to be discriminated) remains constant while their variance is reduced, discrimination sensitivity d' increases.

Improvement of discrimination sensitivity can also be related to familiarity or perceptual learning, which serves to reduce uncertainty. It has been found that discrimination sensitivity can improve through practice (Epstein, 1967; Gibson, 1967). A point-light human walker can be considered to be a well-practiced stimulus because a human walker is extremely familiar. Hence, it should be expected that discrimination sensitivity will be higher for a recognized point-light human figure than for an unrecognizable set of dots. Indeed, in addition to the work by Tadin et al. (2002), it has been found that upright faces and words also yielded better discrimination sensitivity than did inverted ones (Martelli, Majaj, & Pelli, 2005).

For the hand–elbow distance discrimination, because one forearm was shortened and the other forearm lengthened by moving the hand positions with equal distance, the peaks of the elbow posterior estimations were unchanged. However, the peaks of the hand posterior estimations were shifted after recognition in such a way that the shorter forearm was effectively lengthened and the longer forearm shortened by the same amount. This is because the prior distributions of the hands' positions after recognition incorporate the knowledge that the two forearms are usually of equal length. Thus, according to signal detection theory, the two Gaussians were closer after recognition, which would reduce discrimination sensitivity. Yet at the same time their variance was reduced also, which would increase discrimination sensitivity. It turned out that the decrease outweighed the increase; hence, the final outcome was decreased sensitivity.

An intuitive explanation of this outcome is as follows. Imagine, at one extreme, the prior is completely uninformative (the standard deviation is infinitely large). Then each point's posterior estimation of its position is completely determined by the likelihood function. This can be considered as the situation before recognition. At the other extreme, imagine that the prior is extremely certain (being a δ function). In this situation, each point's posterior positional estimation is completely determined by the prior while the likelihood function is irrelevant. Because the prior indicates that the forearms are equal in length, discrimination sensitivity would be zero.

To conclude, our findings that sensitivity in stereoscopic depth discrimination can either increase or decrease after recognizing the stimulus being a human figure fit well with Bayesian analyses of human vision (Gregory, 1970; Kersten & Yuille, 2003; Knill & Richards, 1996; Rock, 1997), according to which prior expectations are integrated with sensory information to optimize perceptual inferences. In the process of shape recognition, the visual system is capable of exploiting information based upon prior knowledge about the geometric structure of shapes. Such information in turn affects the accuracy of stereoscopically defined depth cues, which are often assumed to subserve shape recognition.

Appendix

In this appendix, we derive model predictions of discrimination sensitivity change in the hand–hip and hand–elbow distance discriminations. Two prior probability distributions are assumed for each point's depth coordinate z that is contingent on recognition. Before recognition, the depth coordinate of each point is assumed completely unknown by the prior, which is modeled by a Gaussian with an infinite variance. After recognition, the depth coordinate is assumed more certain, modeled by a Gaussian with a finite variance. In both cases, the likelihood function is assumed to be the same and as a Gaussian function, which is determined by disparity calculation from a point's left- and right-eye images. Given that the 2D projections of two distances are equal, 3D distance discrimination is equivalent to relative depth discrimination.

We first derive the hand–hip distance discrimination, because this is more straightforward.

Increase in discrimination sensitivity

For simplicity, we assume that, before recognition, the prior probability of a point's position in depth is completely uninformative because the stimulus is a meaningless set of points. In other words, we assume that the standard deviation of the prior before recognition is infinity. After recognition, each point becomes a meaningful joint of a human body. There is a prior probability of a point's position in depth (z)

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(z-z_0)^2}{2\sigma_p^2}\right], \tag{A1}$$

where z_0 is the expected depth of the point for a natural posture of a human body. At the same time, the likelihood function, that is, the binocular disparity information δ conditioned on the point's z-coordinate, can be written as

$$p(\delta|z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-\delta)^2}{2\sigma^2}\right]. \tag{A2}$$

After recognition, the prior probability's variance is reduced to be finite. At the same time, because there is no inconsistency between the prior expectation of a point's depth position z_0 and its estimate from binocular disparities δ, $z_0 = \delta$. Hence, the posterior probability distribution of this point's z-position is

$$p(z|\delta) \propto \exp\left[-\frac{(z-\delta)^2}{2(1/\sigma^2 + 1/\sigma_p^2)^{-1}}\right]. \tag{A3}$$

That is, the posterior probability distribution of the point's z-position has a smaller variance, while its peak position is unchanged after recognition. Because the distribution of hand–hip z-distance is the convolution of the two posterior distributions, and because the hand point and hip point have the same posterior distributions in shape, the resultant distribution is also a Gaussian with a variance of $2(1/\sigma^2 + 1/\sigma_p^2)^{-1}$ (recall that in a convolution, variances add up). Let us denote Δz as the difference between the left hand–hip z-distance and the right hand–hip z-distance. Then, discrimination sensitivity after recognition, according to standard signal detection theory in a 2AFC discrimination, is

$$d'_{\text{post}} = \frac{\sqrt{2}|\Delta z|}{\sqrt{2(1/\sigma^2 + 1/\sigma_p^2)^{-1}}} = \frac{|\Delta z|}{\sqrt{(1/\sigma^2 + 1/\sigma_p^2)^{-1}}}. \tag{A4}$$

Because no point's mean z-position is shifted after recognition, Δz is the same before recognition. However, before recognition, $\sigma_p \rightarrow \infty$; therefore, discrimination sensitivity before recognition is

$$d'_{\text{pre}} = \frac{|\Delta z|}{\sigma} < d'_{\text{post}}. \tag{A5}$$

Intuitively, recognition makes the prior distribution more localized and hence more certain. Therefore, the resultant posterior distribution is more certain also, whereas the difference in relative depth remains the same. In other words, the noise is reduced while the signal remains the same, hence d' is increased.

Decrease in discrimination sensitivity

We now show that sensitivity decreases after recognition in the hand–elbow discrimination. The only difference from the condition above is that the prior probability distribution of a hand point no longer centers at the same location as the likelihood function does because after recognition the prior expects the forearms to be equally long. Without loss of generality, we assume that the peak position of the prior for an elbow position still coincides with its likelihood peak, and that only a hand's z-position shifts. The prior or expected location of a hand is such that the two forearms will be equal in length.

Therefore, the posterior estimation of a hand's z-position is

$$p(z|\delta) \propto \exp\left[-\frac{(z-(\sigma^2 z_0 + \sigma_p^2\delta)/(\sigma^2+\sigma_p^2))^2}{2(1/\sigma^2 + 1/\sigma_p^2)^{-1}}\right], \tag{A6}$$

where z_0 is now the peak of the prior after recognition (i.e., the expected position of a hand in a normal body posture).

The maximum a posteriori (MAP) estimate of the hand position, relative to its corresponding elbow position, is $(\sigma^2 z_0 + \sigma_p^2 \delta)/(\sigma^2 + \sigma_p^2)$. Therefore, because one forearm in the stimulus is shorter than expected and the other forearm longer by the same amount, the absolute difference between the lengths of the two forearms is

$$\Delta z = 2\left|z_0 - \frac{\sigma^2 z_0 + \sigma_p^2 \delta}{\sigma^2 + \sigma_p^2}\right| = 2\frac{\sigma_p^2}{\sigma^2 + \sigma_p^2}|z_0 - \delta|, \tag{A7}$$

as compared to $\Delta z = 2|z_0 - \delta|$ before recognition. The discrimination sensitivity d' after recognition is

$$d'_{\text{post}} = \frac{\Delta z}{\sqrt{\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_p^2}\right)^{-1}}} \frac{2|z_0-\delta|}{\sigma}\sqrt{\frac{\sigma_p^2}{\sigma^2+\sigma_p^2}} < d'_{\text{pre}}. \tag{A8}$$

The intuition is less straightforward. We offer two here. The first is that, after recognition, the difference in the distances (i.e., the signal, or the nominator in the d' equation) is a factor of $(\sigma_p^2)/(\sigma^2 + \sigma_p^2) \leq 1$ from before recognition, whereas the standard deviation (the denominator of the d' equation) is the square root of the same factor. Thus, the signal is reduced more than the uncertainty is. The second intuition is that when the prior is completely uncertain, $\sigma_p \rightarrow \infty$, d' is unchanged as compared to before recognition. When the prior is completely certain, $\sigma_p \rightarrow 0$, it overrules the likelihood function, hence the posterior is completely determined by the prior. Because the prior says that the forearms should be equally long, discrimination is at chance, $d' = 0$. When we consider σ_p as decreasing from infinity down to zero, there is no reason for the resultant d' to be nonmonotonic. Hence, when the prior indicates that the difference should be smaller, discrimination sensitivity will decrease no matter how uncertain the prior is.

Acknowledgment

We thank Drs. Pam Lipson and Pawan Sinha for providing the original point-walker stimuli; Drs. David Bennett, Keith Holyoak, Nestor Matthews, and Benjamin Thompson for helpful comments; and Colleen Honigsberg and Tracy Zieve for assistance in conducting the experiments. HL was supported by an International Fellowship from the American Association of University Women.

Commercial relationships: none.
Corresponding author: Zili Liu.
Email: zili@psych.ucla.edu.
Address: Department of Psychology, University of California, Los Angeles, CA, USA.

References

Adams, D. L., & Zeki, S. (2001). Functional organization of macaque V3 for stereoscopic depth. *Journal of Neurophysiology, 86,* 2195–2203. [PubMed] [Article]

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193,* 31–35.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86,* 2054–2068. [PubMed] [Article]

Badcock, D. R., & Schor, C. M. (1985). Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America A, 2,* 1211–1216. [PubMed]

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, 211,* 599–622. [PubMed]

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 1,* 254–257. [PubMed]

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers.

Cooper, F. S., Delattre, P., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America, 24,* 597–606.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394,* 677–680. [PubMed]

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience, 19,* 1398–1415. [PubMed] [Article]

DeAngelis, G. C., & Newsome, W. T. (2004). Perceptual "read-out" of conjoined direction and disparity maps in extrastriate area MT. *PLoS Biology, 2,* E77. [PubMed] [Article]

Epstein, W. (1967). *Varieties of perceptual learning*. New York: McGraw-Hill.

Fahle, M., & Westheimer, G. (1988). Local and global factors in disparity detection of rows of points. *Vision Research, 28,* 171–178. [PubMed]

Farell, B. (1998). Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature, 395,* 689–693. [PubMed]

Feldman, J. (2000). Bias toward regular form in mental shape spaces. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 152–165. [PubMed]

Fisher, G. H. (1968). "Mother, father, and daughter": A three-aspect ambiguous figure. *American Journal of Psychology, 81,* 274–277. [PubMed]

Gibson, E. J. (1967). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, Educational Division, Meredith Corporation.

Glennerster, A., & McKee, S. P. (1999). Bias and sensitivity of stereo judgements in the presence of a slanted reference plane. *Vision Research, 39,* 3057–3069. [PubMed]

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, New York: Robert E. Krieger Publishing Company.

Gregory, R. L. (1970). *The intelligent eye*. London: Weidenfeld and Nicolson.

Gregory, R. L. (1978). *Eye and brain: The psychology of seeing*. New York: McGraw-Hill.

Grenander, U. (1993). *General pattern theory*. New York: Oxford University Press.

Grunewald, A., Bradley, D. C., & Andersen, R. A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *Journal of Neuroscience, 22,* 6195–6207. [PubMed] [Article]

Hartung, B., Schrater, P. R., Bülthoff, H. H., Kersten, D., & Franz, V. H. (2005). Is prior knowledge of object geometry used in visually guided reaching? *Journal of Vision, 5*(6), 504–514, http://journalofvision.org/5/6/2/, doi:10.1167/5.6.2. [PubMed] [Article]

Helmholtz, H. v. (1866/1924). *A treatise on physiological optics* (Vol. 1). New York: Optical Society of America.

Hochberg, J., & Peterson, M. A. (1987). Piecemeal organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General, 116,* 370–380. [PubMed]

Hou, F., Lu, H., Zhou, Y., & Liu, Z. (in press). Amodal completion impairs stereo acuity discrimination. *Vision Research*.

Johanssen, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.

Kersten, D., Bülthoff, H. H., Schwartz, B. L., & Kurtz, K. J. (1991). Interaction between transparency and structure from motion. *Neural Computation, 4,* 573–589.

Kersten, D., Knill, D. C., Mamassian, P., & Bülthoff, I. (1996). Illusory motion from shadows. *Nature, 379,* 31. [PubMed]

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13,* 150–158. [PubMed]

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

Kumar, T., & Glaser, D. A. (1992). Depth discrimination of a line is improved by adding other nearby lines. *Vision Research, 32,* 1667–1676. [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]

Liu, Z., Jacobs, D. W., & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research, 39,* 4244–4257. [PubMed]

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

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research, 41,* 2653–2668. [PubMed]

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1), 58–70, http://journalofvision.org/5/1/6/, doi:10.1167/5.1.6. [PubMed] [Article]

McKee, S. P. (1983).The spatial requirements for fine stereoacuity. *Vision Research, 23,* 191–198. [PubMed]

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779. [PubMed]

Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research, 24,* 1063–1073. [PubMed]

Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67,* 45–71. [PubMed]

Morgan, M. J., & Castet, E. (1997). The aperture problem in stereopsis. *Vision Research, 37,* 2737–2744. [PubMed]

Mussap, A. J., & Levi, D. M. (1995). Amodal completion and vernier acuity: Evidence of 'top-but-not-very-far-down' processes? *Perception, 24,* 1021–1048. [PubMed]

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257,* 1357–1363. [PubMed]

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America, 43,* 907–913. [PubMed]

Peterson, M. A., & Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception & Performance, 9,* 183–193.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. [PubMed]

Rock, I. (1997). *Indirect perception*. Cambridge, MA: MIT Press.

Sherif, M. (1935). A study of some social factors in perception: Chapter 3. *Archives of Psychology, 27,* 23–46.

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1315–1341. [PubMed]

Tadin, D., Lappin, J. S., Blake, R., & Grossman, E. D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience, 5,* 1010–1015. [PubMed]

Vreven, D., McKee, S. P., & Verghese, P. (2002). Contour completion through depth interferes with stereoacuity. *Vision Research, 42,* 2153–2162. [PubMed]

Wallace, J. M., & Mamassian, P. (2004). The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces. *Vision Research, 44,* 2253–2267. [PubMed]

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5,* 598–604. [PubMed]

Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27,* 1361–1368. [PubMed]

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research, 40,* 1969–1978. [PubMed]