

Integrating multiple cues to depth order at object boundaries

Ellen C. Hildreth · Constance S. Royden

© Psychonomic Society, Inc. 2011

Abstract We examined the interaction between motion and stereo cues to depth order along object boundaries. Relative depth was conveyed by a change in the speed of image motion across a boundary (motion parallax), the disappearance of features on a surface moving behind an occluding object (motion occlusion), or a difference in the stereo disparity of adjacent surfaces. We compared the perceived depth orders for different combinations of cues, incorporating conditions with conflicting depth orders and conditions with varying reliability of the individual cues. We observed large differences in performance between subjects, ranging from those whose depth order judgments were driven largely by the stereo disparity cues to those whose judgments were dominated by motion occlusion. The relative strength of these cues influenced individual subjects' behavior in conditions of cue conflict and reduced reliability.

Keywords 3-D perception · Cue integration · Binocular vision · Stereopsis

One of the challenging tasks of early visual processing is segmentation of the image into regions corresponding to distinct objects. To accomplish this, the visual system must

first locate potential object boundaries and infer the depth order of surfaces meeting at these boundaries. Such information can facilitate grouping processes that organize boundary fragments into extended objects, and processes that recover the three-dimensional shape of object surfaces from the two-dimensional image. Many visual cues, such as color, texture, motion, and stereo disparity, contribute to the task of detecting object boundaries. In this study, we explored the contributions of motion and stereo cues to the assessment of depth order at object boundaries.

Image motion and stereo disparity provide multiple cues to the relative depth of two surfaces meeting at an edge in the image. When an observer translates relative to a stationary scene, features on surfaces closer to the observer move with higher image speeds than those on surfaces that are farther away (Gibson, 1966; Howard & Rogers, 2002; Longuet-Higgins & Prazdny, 1981). In addition, features on a more distant surface appear or disappear as the surface moves behind an occluding object. The border between the two surfaces moves with the closer, occluding surface (Feldman & Weinshall, 2008; Gibson, Kaplan, Reynolds, & Wheeler, 1969; Kaplan, 1969; Proffitt, Bertenthal, & Roberts, 1984; Royden, Baker, & Allman, 1988; Thompson, Mutch, & Berzins, 1985; Yonas, Craton, & Thompson, 1987). Under binocular viewing, the difference in stereo disparity between adjacent image regions provides a direct cue to the depth order of the two surfaces. The presence of half-occlusions, surface regions seen by only one eye, also reveals a surface that is located behind an adjacent occluding object (Anderson & Nakayama, 1994; Egnal & Wildes, 2002; Harris & Wilcox, 2009; Howard & Rogers, 2002). The ability to infer depth from these regions of half-occlusion is referred to as *da Vinci stereopsis*. Many of these motion and stereo cues can be analyzed early in the processing of the retinal image to detect the location of object boundaries prior

E. C. Hildreth (✉)
Department of Computer Science, Wellesley College,
106 Central Street,
Wellesley, MA 02481, USA
e-mail: ehildreth@wellesley.edu

C. S. Royden
Department of Mathematics and Computer Science,
College of the Holy Cross,
Worcester, MA, USA

to the computation of a detailed representation of the 3-D structure and motion of objects in the scene.

Motion and stereo cues often provide clear and consistent information about the depth order of surfaces at a boundary and can reinforce one another. Situations arise, however, when one or more cues provide ambiguous or unreliable information about depth order. For example, relative image speed is not always a reliable cue to depth order. An independently moving object at a greater distance from the viewer can move in a way that yields faster image speeds relative to an adjacent surface that is closer to the viewer. Rapid eye rotation during pursuit adds image motion that can mask velocity differences due to changes in depth (van den Berg & Beintema, 2000). Stereo disparity can be unreliable due to the limited range of disparity that can be fused at a single fixation distance. When there is a large change in depth between adjacent surfaces, the stereo disparity of one or both surfaces may be outside the range that can be fused by the stereo system. There is evidence that observers can sense depth directly in diplopic images (Wilcox & Allison, 2009; Ziegler & Hess, 1997), although the impression is less compelling than that obtained from fused stereograms. As distance increases, larger differences in depth are needed to detect depth changes and infer depth order from stereo disparity (Cutting & Vishton, 1995), although the presence of small stereoacuity thresholds near the fovea suggests that stereo may provide a reliable depth order cue over fairly large distances for centrally viewed surfaces (Tyler, 2004). When surfaces lack visual texture in the vicinity of a border, both stereo and motion cues to depth order can be weak (Yonas et al., 1987). Finally, motion and stereo cues sometimes provide conflicting information about depth changes. For example, when viewing a computer display of a dynamic 3-D scene with both eyes, image motion conveys 3-D depth and movement, while the stereo system indicates a flat 2-D scene. In these scenarios, the question arises, how do we integrate multiple cues to depth order at an edge in situations in which one or more cues provide weak, ambiguous, or conflicting information?

Physiological studies reveal neurons as early as area V2 in monkey cortex that respond to boundaries defined by changes in image motion, stereo disparity, or luminance (Bredfeldt & Cumming, 2006; Marcar, Raigel, Xiao, & Orban, 2000; Qiu & von der Heydt, 2005; von der Heydt, Qiu, & He, 2003; von der Heydt, Zhou, & Friedman, 2000; Zhou, Friedman, & von der Heydt, 2000; for a review, see Orban, 2008). Sometimes these neurons encode the depth order of surfaces meeting at an edge (Qiu & von der Heydt, 2005; von der Heydt et al., 2000; Zhou et al., 2000). Cue-invariant boundary representations are also found in the responses of neurons in area V4 (Mysore, Vögels, Raigel, & Orban, 2006) and the inferotemporal visual area (Liu,

Vögels, & Orban, 2004). The pervasiveness of center-surround mechanisms for motion and stereo processing in area MT suggests that it may play a role in object segmentation (Born & Bradley, 2005; Bradley & Andersen, 1998; Tadin & Lappin, 2005), although MT cells do not exhibit the same selectivity for the position and orientation of motion-defined boundaries as found in area V2 (Marcar, Xiao, Raigel, Maes, & Orban, 1995). Addressing the question of how boundary cues are combined perceptually may shed light on the neural integration of motion and stereo in the analysis of depth changes across surface boundaries.

Broadly speaking, information about depth order along a surface boundary can be combined across multiple cues in several ways. Two cues providing consistent information may reinforce or enhance one another. In this situation, if one or more of the cues is degraded, their combination may result in a stronger, more consistent percept of the location of the boundary and change in depth across the boundary, relative to the percept obtained from each cue in isolation (Rivest & Cavanagh, 1996). If one cue yields an ambiguous depth order, a second cue may resolve this ambiguity (Pong, Kenner, & Otis, 1990). When two cues provide conflicting information, one cue may dominate. For example, observers may perceive the depth order indicated by stereo disparity, even when motion occlusion along the surface border indicates the opposite depth order. Finally, the judgment of depth order could reflect a compromise between cues, such as a weighted combination of depth information provided by different cues in the vicinity of a boundary. The weights associated with individual cues may depend on the reliability, consistency, or relevance of the cues (Bruno & Cutting, 1988; Jacobs, 2002; Landy, Maloney, Johnston, & Young, 1995; Yuille & Bülthoff, 1996).

Recent studies have drawn attention to significant individual differences in the processing of stereo and motion information (e.g., Nefs, O'Hare, & Harris, 2010; Wilmer, 2008; Wismeijer, Erkelens, van Ee, & Wexler, 2010). Such differences can be exploited in the study of the functional organization and utility of visual processing (Wilmer, 2008). If individual differences exist in the processing of boundary information from stereo and motion cues in isolation, these differences may impact the strategies used for cue integration. For example, if a subject has substantially better performance with one of these cues relative to the other, the stronger cue may be more likely to dominate in cue conflict conditions, and there may be less enhancement in performance when consistent stereo and motion cues are combined.

We present the results of experiments that explored the interaction between motion and stereo cues in the perception of the depth order of adjacent frontoparallel surfaces moving horizontally or toward the observer. The separation

in depth between the surfaces was conveyed by the relative speed of movement of image features on either side of the surface boundary (motion parallax, or relative speed cue); the disappearance of features along a border when one surface moved behind an occluding object (motion occlusion); and a difference in the stereo disparity of features on the two surfaces. We first compared the perception of depth order from relative speed alone to that resulting from the addition of motion occlusion (Exp. 1) or a difference in stereo disparity (Exp. 2). These two experiments also explored how depth order influences the perception of relative movement in depth of the two surfaces. We then examined how conflicts are resolved when motion occlusion and stereo disparity convey opposite depth orders (Exp. 3) and how the integration of the two cues depends on the strength or reliability of the cues (Exp. 3 and 4). Finally, we discuss important differences in performance across subjects and their implications for the ways in which stereo and motion cues are combined in the analysis of depth order at object boundaries.

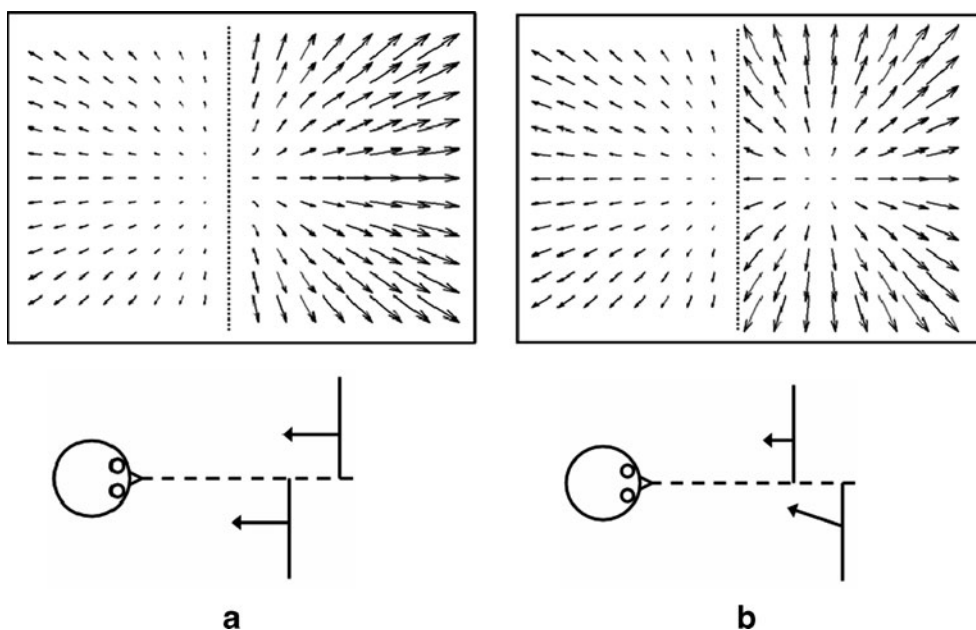
Experiment 1

In this experiment, we examined the interaction between depth order constraints from relative image speed and motion occlusion. The visual displays simulated two frontoparallel surfaces, located side by side and moving toward the viewer, with a vertical border between the two surfaces centered in the image. Patterns of image motion created in two sample scenarios are shown in Fig. 1. In the case of Fig. 1a, both surfaces move in depth only, creating

an expanding pattern of image motion with a single focus of expansion at the center. The simulated depth and speed of movement in depth of the right surface results in higher image speeds, providing a cue to the relative depth of the surfaces. In the case of Fig. 1b, an additional leftward translation of the right surface shifts the focus of expansion associated with this surface to the right. The boundary between the two surfaces remains stationary in the image over time, as if fixed to the left surface, and features on the right surface disappear when they reach the boundary. The resulting motion occlusion cue indicates that the right surface is sliding behind the left surface. If the observer assumes that a faster-moving surface is closer in depth, as expected from motion parallax, this would lead to a contradiction. This experiment compared the percept of depth order when no motion occlusion was present (Fig. 1a) to the case in which motion occlusion was added (Fig. 1b). Of particular interest was the situation in which motion occlusion and relative image speed provided conflicting cues to depth order.

Ono, Rogers, Ohmi, and Ono (1988) examined the interaction between conflicting motion parallax and occlusion cues to depth order for random-dot surfaces in the case in which the relative movement between the viewer and surfaces was in the lateral direction. Depth order judgments were dominated by motion occlusion, except in the case in which the simulated depth difference between the surfaces was very small, yielding few appearing or disappearing dots along the surface border. Braunstein, Andersen, and Riefer (1982) found that motion occlusion dominates depth order judgments in the perception of 3-D structure from a rotating sphere of random dots. We expect from these

Fig. 1 (a) Image velocity field resulting from two frontoparallel planes at different depths translating toward the viewer, with a focus of expansion at the center of the image. Below the image velocity field is a birds' eye view of one 3-D configuration of surfaces and motion that would yield this pattern of motion. (b) Similar to panel a, except that the plane on the right also translates slightly to the left, creating a second focus of expansion on the right side of the image. In the actual displays, moving dots appeared at random locations in the image, and there was no explicit marking of the central border



observations that depth order judgments for surfaces moving toward the observer will also be dominated by motion occlusion cues.

In the scenarios depicted in Fig. 1, relative image speed alone provides an ambiguous cue to depth order. In the case of Fig. 1a, the right surface could be located at a greater distance from the observer and be moving with a substantially greater speed toward the observer that yields faster image velocities. If the left and right surfaces are located at depths Z_l and Z_r , and move toward the viewer with constant speeds T_{zl} and T_{zr} , then under perspective projection, the image velocities (v_{xl}, v_{yl}) and (v_{xr}, v_{yr}) at locations (x, y) are given by

$$(v_{xl}, v_{yl}) = (x, y) * (T_{zl}/Z_l), \quad (v_{xr}, v_{yr}) = (x, y) * (T_{zr}/Z_r)$$

The ratio T_z/Z is the inverse of the time-to-contact (TTC) of an approaching surface. Because image velocity depends on this ratio rather than on T_z or Z alone, the direct relationship between Z_l and Z_r cannot be determined from the image velocities. One approach to inferring depth order for the scenario depicted in Fig. 1a is to assume that the two surfaces together move in a rigid configuration. In this case, $T_{zl} = T_{zr}$, so the surface with higher image speeds is closer to the observer, and the ratio of depths that gives rise to the observed image velocities is equal to the ratio of the TTC for the two surfaces: $Z_l/Z_r = \text{TTC}_l/\text{TTC}_r$. For the particular scenario illustrated in Fig. 1b, there is no rigid configuration of moving surfaces consistent with the image motions, so the rigidity assumption alone is not sufficient to determine depth order.

Hildreth, Ando, Andersen, and Treue (1995) proposed a model for computing 3-D structure from motion in which the 3-D shape emerges over time through incremental changes. At each moment, the model first computes a new 3-D structure that is consistent with image motion and maximizes the rigidity of the evolving structure (Grzywacz & Hildreth, 1987; Ullman, 1984). A smooth surface is then computed from this skeleton 3-D structure, incorporating information about boundaries from motion discontinuities and other depth cues. The reconstructed surface specifies a new depth for image features that is used to compute their 3-D structure at the next moment in time. In the absence of other 3-D cues, the model initializes the structure to be flat, with all surface features at the same depth. In the case of two frontoparallel surfaces directly approaching the observer (Fig. 1a), the surfaces are initially placed at the same depth Z_{init} —for example, $Z_l = Z_r = Z_{\text{init}}$. The computed 3-D structure consists of two surfaces moving rigidly toward the viewer, with speeds of movement in depth consistent with the image velocities. The ratio of these speeds is equal to the ratio of the TTC for the two surfaces: $T_{zl}/T_{zr} = \text{TTC}_r/\text{TTC}_l$. Before the motion discontinuity along the central

border becomes apparent, the model interpolates a smooth surface across the border region. Once the boundary is detected, the surface reconstruction process yields two flat planes with a depth discontinuity at the center. Given this model, we would predict that the surfaces will appear at the same depth initially, and the surface with a smaller TTC will appear to move faster in depth, eventually moving out in front of the other surface. For the scenario shown in Fig. 1b, when the motion occlusion becomes apparent, it can force a particular depth order for the two surfaces along the central border. The ongoing surface reconstruction process in the Hildreth et al. (1995) model can incorporate this depth order constraint when computing a new surface representation at each moment. For the particular example shown in Fig. 1b, the application of the model yields the prediction that the right surface will appear to remain behind the left surface, moving faster in depth than the left surface. The computed 3-D structure of the right surface would not be strictly rigid, and would expand slightly over time.

Both approaches described above yield the prediction that in the absence of an explicit depth order cue from motion occlusion, the surface with features moving at higher image speeds will appear to be in front. If we assume that the surfaces move in a rigid configuration, the depth difference between the surfaces should appear to remain constant as the surfaces move toward the observer. If, on the other hand, we assume that the surfaces initially appear at the same depth and can move independently in a way that maximizes the rigidity of each surface, then the surfaces should appear to separate in depth over time. When motion occlusion is present, it should provide an unambiguous cue to depth order. As a consequence, we predict that the motion occlusion cue will determine perceived depth order. Depth order should also influence the percept of the relative movement in depth between the two surfaces. For example, if the motion occlusion cue indicates that the surface with faster image speeds is in back, as is shown in Fig. 1b, the back surface should appear to move faster in depth and approach the front surface over time.

Method

Subjects A group of 19 paid observers with normal or corrected-to-normal vision participated in this experiment. All were naïve about the purpose of the experiment.

Stimuli and procedure The visual displays simulated two frontoparallel planes of random dots, located side by side. A vertical border between the surfaces remained stationary at the center of the image. The border was defined only by relative motion. Sequences of images were created by

simulating movement of the surfaces toward the plane of the viewer, using different combinations of initial depths and speeds of movement in the horizontal and depth directions. The stimuli were generated on a Power Mac G4 using the Psychophysics Toolbox software (Brainard, 1997) and presented on a LaCie 19-in. CRT monitor. The display had a resolution of $1,024 \times 768$ pixels and a refresh rate of 85 Hz. The displays contained red dots on a black background and were viewed monocularly in the dark. The dots subtended $9'$ of visual arc, and their size remained constant over time. The image of the surfaces subtended $20^\circ \times 20^\circ$ of visual arc from a viewing distance of 0.4 m. The outer borders of the image remained stationary over time. There were an average of 4,000 dots per frame, giving an average dot density of 10 dots/deg², and dots disappeared at the outer borders of the image as the surfaces approached the observer. A white fixation cross at the center of the display remained stationary for the duration of the stimulus, which lasted 0.5 s. The stimulus contained 10 frames, with 50 ms per frame and no interframe interval (the frame rate matched that used in Exps. 2–4, which was limited by software constraints for presenting the stereo images). For each trial, the fixation cross appeared and the subject pressed a key to display the random-dot pattern, which was immediately in motion. The display disappeared as soon as the movement stopped, and the subject responded with a second keypress.

The combinations of simulated initial depth and movement in depth used to generate the image motions corresponded to seven values of the initial TTC (ratio Z/T_z): 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, and 3.0 s (the image velocities are the same for a given TTC, regardless of the particular Z and T_z values used to generate them). The experiment consisted of three blocks, each lasting 4–5 min. In the first block, the surfaces moved in depth only, as shown in Fig. 1a. There were 16 combinations of movement for the two surfaces, with differences in initial TTC between the surfaces of 0.25, 0.5, 0.75, and 1.0 s. The surfaces moved with constant speed, so this difference remained constant over the trial. There were eight conditions with the smaller TTC on the left, and eight conditions with the smaller TTC on the right. Half of the conditions had a smaller TTC of 1.5 s (paired with a surface having a TTC of 1.75, 2.0, 2.25, or 2.5 s), and half had a smaller TTC of 2.0 s (paired with a surface having a TTC of 2.25, 2.5, 2.75, or 3.0 s) at the start of the trial. This block had 10 repetitions of the 16 conditions, randomly ordered. Observers indicated with a keypress which surface appeared to be in front.

In the second and third blocks, one surface translated in depth only, while the other surface had a small horizontal translation toward the center of the image, as shown in Fig. 1b. The border between the surfaces remained

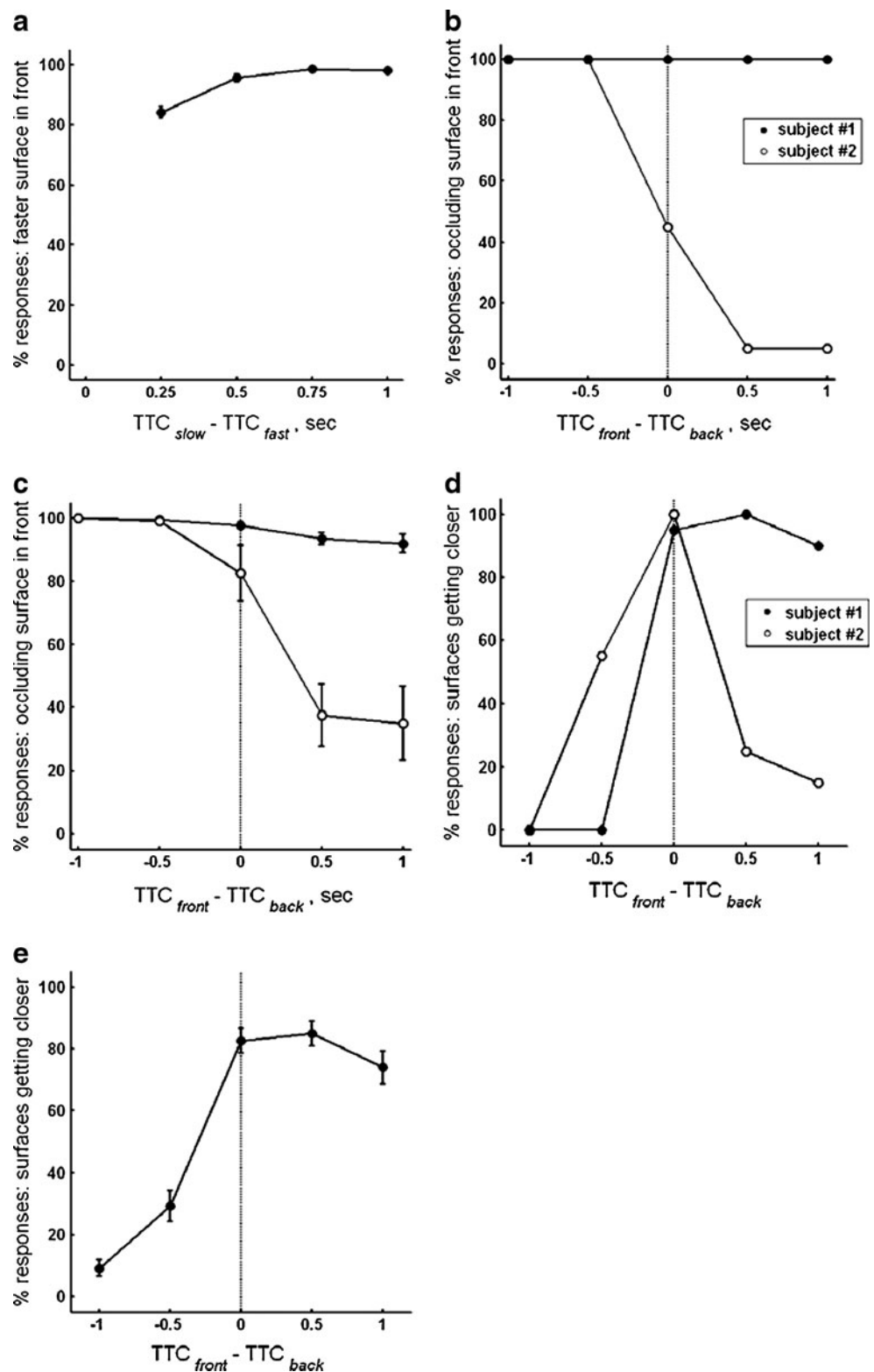
stationary at the center of the image, and dots on the horizontally shifting surface disappeared when they reached the border, creating the impression that this surface was being occluded. A high dot density was used (10 dots/deg²) in order to provide a strong occlusion cue. The simulated motion parameters resulted in the occluded surface having faster image motions in half of the trials, and the occluding surface moving faster in the other half. The initial TTC for each surface was 1.5, 2.0, or 2.5 s. There were five values for the difference in TTC between the two surfaces, $TTC_{\text{front}} - TTC_{\text{back}}$, where TTC_{front} refers to the initial TTC of the simulated front, occluding surface, and TTC_{back} refers to the TTC of the simulated back, occluded surface. The five differences were -1.0 , -0.5 , 0.0 , 0.5 , and 1.0 s. For the difference of 0.0 , both initial TTC's were 2.0 s, and for the differences of -0.5 and 0.5 , the initial TTC's of the two surfaces were 1.5 and 2.0 s. The horizontal component of motion of the occluded surface was chosen so that the extent of the horizontal displacement of the occluded surface at the border between the surfaces was the same for all movements in depth. The corresponding 3-D direction of motion of the occluded surface was 2° , 3° , or 4° to the left or right of the direction toward the viewer. This 3-D direction of motion is illustrated in the birds' eye view shown in Fig. 1b, where the right surface is occluded. The second and third blocks each contained 10 repetitions of 10 conditions (five values for the TTC difference, with the occluded surface on the left or right).

In the second block, subjects judged which surface appeared to be in front on each trial, while in the third block, they judged whether the surfaces appeared to become closer together or farther apart in depth as they moved toward the observer. If the motion occlusion cue imposes a strict constraint on the order of the surfaces in depth that cannot be overridden by other cues, and the surface that appears in back (the occluded surface) has a faster simulated motion relative to the front surface, then the surfaces should appear to move closer together in depth over time. If the back surface has a slower simulated motion, the surfaces should appear to move farther apart.

Results

The results for the first block, in which the visual displays provided only the relative speed cue, are summarized in Fig. 2a. The percentages of trials on which the surface containing dots with faster image speeds was judged to be in front are plotted against the difference in TTC between the two surfaces, $TTC_{\text{slow}} - TTC_{\text{fast}}$, where TTC_{slow} and TTC_{fast} refer to the TTC for the surfaces with slower and faster dot speeds, respectively. This measure remains constant as the surfaces approach the viewer. The graph shows mean data for the 19 subjects, combining conditions

Fig. 2 Results for Experiment 1. (a) Mean data for the first block. The percentages of trials on which the surface with faster image speeds was judged to be in front are plotted against the difference in the initial time-to-contact (TTC) of the two surfaces, $TTC_{slow} - TTC_{fast}$. (b) Individual data from the second block are shown for 2 subjects, with filled and open circles. The percentages of trials on which the occluding surface was judged to be in front are plotted against the difference in the initial TTC between the surfaces, $TTC_{front} - TTC_{back}$. (c) Mean data for the second block are shown for two subject groups: those whose judgments were highly consistent with the motion occlusion cue (filled circles), and the remaining 6 subjects (open circles). (d) Individual data for the third block, for the same 2 subjects shown in panel b. The percentages of trials on which subjects judged the surfaces to be getting closer together in depth are plotted against the difference in the initial TTC of the two surfaces, $TTC_{front} - TTC_{back}$. (e) Mean data for all subjects for the third block. All error bars show standard errors



with faster dot speeds on the left and right. The data confirm that the surface with faster dot speeds is perceived to be in front. We asked a few subjects to describe their qualitative impression of how the two surfaces moved relative to one another. They reported that, qualitatively, the

surfaces appeared to move independently—they did not appear to move as a rigid configuration. The perceived separation in depth emerged over a short time after the dots appeared and began to move, and the surface with dots moving at higher image speeds appeared to move faster in

depth. As a consequence, the surfaces always appeared to move farther apart in depth over time.

The results of the second block, in which motion occlusion was added and subjects judged which surface appeared in front, are shown in Fig. 2b and c. The percentages of trials on which the simulated occluding surface was judged to be in front are plotted as a function of the difference in TTC between the occluding (front) and occluded (back) surfaces, $TTC_{\text{front}} - TTC_{\text{back}}$. A positive value of this difference corresponds to conditions in which the front, occluding surface had dots with slower image speeds, while a negative value means that the front, occluding surface had faster dot speeds. A vertical line is drawn at the point where $TTC_{\text{front}} - TTC_{\text{back}} = 0$, which represents the transition between conditions in which relative speed and motion occlusion provided consistent cues to depth order ($TTC_{\text{front}} - TTC_{\text{back}} < 0$) or conflicting cues ($TTC_{\text{front}} - TTC_{\text{back}} > 0$). Data for conditions with the occluding surface on the left and right were similar and are combined. Individual data for 2 subjects illustrating the range of behavior observed are shown in Fig. 2b. Subject 1 (filled circles) always judged the simulated occluding surface to be in front. Most subjects (13/19) behaved in this way. Informal reports from a few of these subjects suggest that the surface with faster image motions appeared to be in front for a brief moment at the start of the trial, and the perceived depth order flipped once the occlusion became apparent. Subject 2 (open circles) always judged the surface with faster dot speeds to be in front, ignoring the motion occlusion cue.

Mean data for this block are shown in Fig. 2c. To highlight the differences in performance, subjects were divided into two groups based on their performance in the cue-conflict conditions ($TTC_{\text{front}} - TTC_{\text{back}} > 0$). Filled circles show mean data for subjects with at least 75% of their judgments consistent with the motion occlusion cue for each of these conditions (13 subjects). The high mean values and small error bars for this group indicate a strong preference for judging depth order based on the motion occlusion cue. Open circles show mean data for the remaining 6 subjects, whose performance was either close to chance or strongly favored the speed cue.

Data for the third block, in which subjects judged whether the surfaces appeared to become closer together or farther apart in depth as they moved forward, are shown in Fig. 2d and e. The percentages of trials on which the surfaces were judged to move closer together in depth are plotted as a function of the TTC difference, $TTC_{\text{front}} - TTC_{\text{back}}$, where “front” and “back” refer to the depth order indicated by the simulated motion occlusion. Individual data are shown in Fig. 2d for the same 2 subjects whose data are shown in Fig. 2b. Mean data for all 19 subjects are shown in Fig. 2e. For most subjects, when the occluded

(back) surface had a smaller initial TTC ($TTC_{\text{front}} - TTC_{\text{back}} > 0$), resulting in faster image speeds, the surfaces appeared to get closer together in depth over time. When the occluded surface had a larger TTC, resulting in slower image speeds, the surfaces appeared to separate in depth. This would be expected for subjects whose depth order judgments are dominated by the motion occlusion cue (Subject 1 in Fig. 2d). Those whose depth order judgments are dominated by the relative speed cue should always perceive the surfaces as separating in depth as they approach the observer. This is shown in the data for Subject 2 in Fig. 2d. Whenever the TTC difference was sufficiently large, this subject perceived the surfaces as separating in depth. For all subjects, there appears to be a bias toward perceiving the surfaces as getting closer in depth when they have the same initial TTC ($TTC_{\text{front}} - TTC_{\text{back}} = 0$). A one-way ANOVA combining data for all 19 subjects showed a significant main effect of TTC difference, $F(4, 90) = 60.5, p < .001$.

Discussion

When relative image speed was the only cue to depth order in the displays, the surface with faster image speeds appeared closer to the observer, as expected from motion parallax. Furthermore, for most subjects, motion occlusion provided an especially strong cue to depth order that could override the depth order indicated by the relative speed cue when the two cues were in conflict. This was not surprising, as relative image speed is a less reliable cue to depth order in the general case in which the environment contains independently moving objects. For some subjects (6/19), relative speed also provided a strong cue to depth order that in some cases could override the motion occlusion cue (e.g., Subject 2 in Fig. 2b). The qualitative impressions of surface motions solicited from a few subjects are consistent with those expected from the structure-from-motion model proposed by Hildreth et al. (1995), but further experiments are needed to confirm this.

The depth order indicated by occlusion also influences the perceived relative movement of surfaces in depth. Qualitatively, the surface with faster image motions appears to move faster in depth. For most subjects, when the motion occlusion cue indicates that the faster-moving surface is in back, the two surfaces appeared to move closer together in depth as they moved forward, as if the occluded surface was catching up to the front surface. Informal reports from a few subjects suggested that the perceived depth difference resulting from motion occlusion was very small, but for most subjects, the faster-moving surface did not appear to overtake or pass the front, slower-moving surface over the short, 0.5-s duration of the movement. For these subjects, the depth order constraint imposed by the motion occlusion

cue was fairly strict and not easily overridden by the relative speed cue. These results reinforce the conclusions of Braunstein et al. (1982) and Ono et al. (1988). For a small number of subjects, the motion occlusion cue was not as strong, and both depth order and relative movement in depth were strongly influenced by relative speed.

Experiment 2

This experiment examined the interaction between the depth order constraints from relative image speed and stereo disparity. The visual displays again simulated two frontoparallel surfaces, located side by side and moving toward the viewer, with the vertical border between the two surfaces centered in the image. The simulated depths and movement in depth of the two surfaces resulted in differences in image speed that conveyed differences in depth when viewed monocularly. Stereo disparity was added to the two surfaces and indicated a depth order that was the same as or opposite to the order implied by the relative image speed according to motion parallax. We examined how relative speed and stereo disparity cues were combined in the perception of depth order and the relative movement in depth of the two surfaces.

Method

Subjects A group of 17 naïve subjects with normal or corrected-to-normal vision participated in this experiment. All of these subjects had participated in [Experiment 1](#).

Stimuli and procedure Binocular versions of the moving random-dot patterns were viewed with CrystalEyes stereo shutter glasses made by Stereographics, Inc. The absolute stereo disparity of the surface dots remained constant over time, while their image motion simulated movement of the surfaces in depth. To add stereo disparity, each dot in the original monocular image was copied in the left and right stereo views at locations that were shifted horizontally by half of the specified disparity, on either side of the original dot position. The dots in the left and right views moved with the same direction and speed as the original monocular dot.

This experiment was divided into four blocks. In all blocks, the left and right images each contained an average of 2,000 dots per frame subtending 12' of visual arc, displayed as red dots on a black background. The overall image subtended $20^\circ \times 20^\circ$ of visual arc when viewed from a distance of 0.4 m. This resulted in an average dot density of 5 dots/deg². In the first block, subjects performed a static stereo task in which two frontoparallel surfaces of dots, arranged side by side, were displayed with different

disparities that were both crossed or both uncrossed. This block tested whether subjects had adequate stereo vision for subsequent experiments. There were four combinations of disparity for the two surfaces—(12', 18'), (12', 24'), (-12', -18'), and (-12', -24')—and the front surface appeared on the left or the right in equal numbers of trials. The session contained 10 repetitions of the eight conditions. In each trial, a white fixation cross at zero disparity appeared at the center of the display, and after a keypress, the static stereo pattern appeared for 0.5 s (this duration is sufficient for most observers to perceive depth from stereo; Patterson et al., 1995). Subjects indicated with a second keypress whether the left or right surface appeared in front.

In the remaining three blocks, the two surfaces moved in depth toward the viewer. Disparity remained constant over time, so the stereo cue indicated that the surfaces were stationary in space, with one surface in front of the other. As before, the motion cues indicated that the surfaces were moving toward the viewer, and the ratio of the initial depth and speed of movement in depth (TTC) was varied across trials. The stimulus duration was again 0.5 s. The central border between the two surfaces was defined only by the relative disparity and movement of the surfaces. This border remained stationary over time. A stationary fixation cross at zero disparity was displayed in the center of the image. The second block contained 10 repetitions of 16 conditions. There were two differences in TTC between the surfaces (0.5 and 1.0 s, with one surface having an initial TTC of 1.5 s and the other having an initial TTC of 2.0 or 2.5 s) and two combinations of stereo disparity, (12', 24') or (-12', -24'). The surface with smaller disparity, and the surface with faster dot speeds, appeared on the left or the right on equal numbers of trials. On each trial, the fixation cross appeared first, and after a keypress the moving stereo pattern appeared for 0.5 s. Subjects indicated with a second keypress whether the left or right surface appeared to be in front. Subjects were asked to maintain fixation throughout the trial, but eye movements were not recorded, so we cannot be certain that fixation was maintained.

The final two blocks included 28 combinations of motion and stereo parameters, divided between the two blocks. There were two combinations of stereo disparity, (12', 24') or (-12', 24'), with the surface with smaller disparity appearing on the left or the right. There were seven combinations of depth and movement parameters, corresponding to seven values for the difference $TTC_{\text{front}} - TTC_{\text{back}}$, where “front” and “back” now refer to the depth order indicated by stereo disparity. One surface had an initial TTC of 1.5 s, while the other had an initial TTC of 1.5, 2.0, 2.5, or 3.0 s, yielding $TTC_{\text{front}} - TTC_{\text{back}}$ values of -1.5, -1.0, 0.0, 1.0, and 1.5. Ten repetitions were presented for each condition. Subjects judged whether the

surfaces appeared to become closer together or farther apart in depth as they moved toward the viewer. In this case, both the fixation cross and the first frame of the movie appeared first, and the subject could take as much time as needed to fuse the stereo pattern while maintaining fixation on the central cross, before making a keypress to initiate the movement of the dots. As a consequence, the initial depth order of the surfaces was determined by the stereo cue. The dots again disappeared when they stopped moving, and the subject indicated a judgment of the relative change in depth (surfaces getting closer together or farther apart in depth) with a second keypress.

Results

All subjects performed above the 95% level on the static stereo task. Figure 3a shows mean results for the second block of this experiment, in which subjects judged which surface appeared in front in displays that combined relative speed and stereo disparity cues to depth order. The percentages of trials on which subjects' judgments agreed with the depth order specified by stereo disparity are plotted against the difference $TTC_{front} - TTC_{back}$, which remained constant throughout each trial. Data for crossed and uncrossed disparities, and for conditions with the front surface appearing on the left and the right, were similar and are combined. Two subjects showed some influence of relative speed in their depth judgments (relative speed had also been a strong cue for these subjects in Exp. 1), but overall, subjects' judgments were dominated by the stereo cue, so that the mean data in Fig. 3a combine the results for all 17 subjects.

Figure 3b shows mean data for the final two blocks, in which subjects judged whether the surfaces appeared to move closer together or farther apart in depth as they moved toward the viewer. Mean data for all 17 subjects are combined. In general, when the front surface, as specified

by stereo disparity, was moving slower, as specified by the motion cue ($TTC_{front} - TTC_{back} > 0$), the surfaces appeared to become closer in depth over time. Otherwise, the surfaces appeared to separate in depth over time. A one-way ANOVA showed a significant main effect of TTC difference, $F(6, 112) = 217.99, p < .001$.

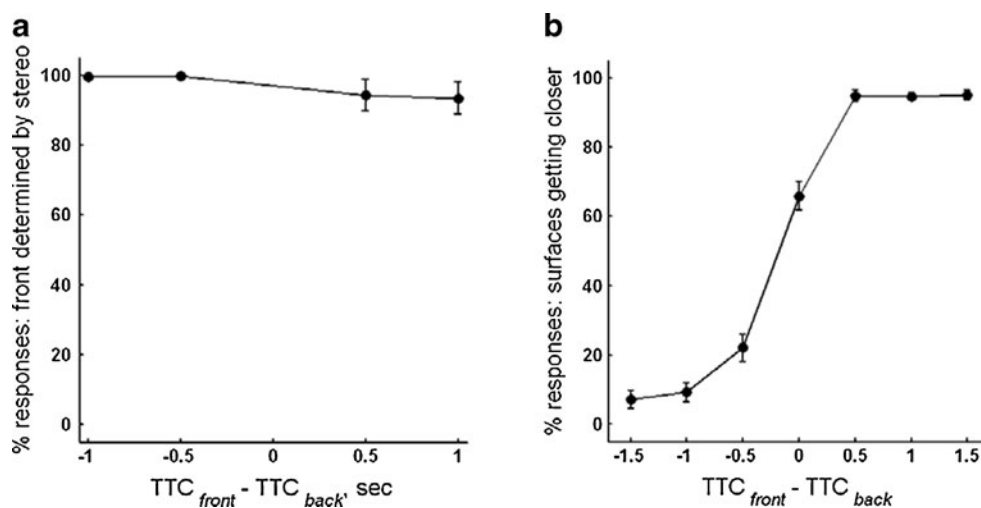
Discussion

For most subjects, stereo disparity provided a strong cue to depth order that could not be overridden by the relative speed cue. The use of displays with a lower dot density in this experiment might have weakened the relative speed cue somewhat, although subjects reported that the individual left or right views, when viewed monocularly, created a strong impression of a depth difference between the surfaces. Over the short extent of the motion, the perceived difference in depth between the two surfaces changed, but their ordering in depth, as specified by the stereo cue, was never violated. Informal observations have suggested that even with a longer viewing time of 1 s, a faster-moving surface in back does not appear to pass the forward surface; perceptually, it remains behind the surrounding surface, in agreement with the stereo cue. The perceived movement of the surfaces in depth is strongly influenced by relative image speed. The stereo cue indicates that the difference in depth between the surfaces remains constant over time, but the difference in image speeds creates the impression that the surfaces are moving closer together or farther apart in depth over time.

Experiment 3

The previous two experiments showed that motion occlusion and stereo disparity provide strong cues to depth order at surface boundaries that, for most subjects, can override

Fig. 3 Results for Experiment 2. The percentages of trials on which judgments of which surface was in front were consistent with the stereo cue (panel a) or trials on which the surfaces were judged to be getting closer (panel b) are plotted against the difference in the initial TTC for the two surfaces, $TTC_{front} - TTC_{back}$. (a) Mean data for the second block, in which subjects' performance was dominated by the stereo cue. (b) Mean data for the final two blocks. All error bars indicate standard errors



inferences about depth order drawn from the analysis of relative image speeds. This experiment addressed the direct interaction between stereo disparity and motion occlusion cues. One aim of this experiment was to test whether two cues that provide weak but consistent information about depth order reinforce one another and yield better performance than either cue alone. Rivest and Cavanagh (1996) found, for example, that for the task of localizing the position of a boundary defined by luminance, color, motion, or texture cues, precision improves when cues are combined. A second aim of this experiment was to examine whether significant differences exist in the ways that individual subjects combine stereo and motion occlusion cues to depth order at object boundaries. We included conditions with only a single cue present, in order to test whether differences in cue combination behavior could be a consequence of differences in subjects' ability to process each cue in isolation.

Method

Subjects A group of 16 naïve subjects with normal or corrected-to-normal vision performed this experiment. All of these subjects had participated in Experiments 1 and 2.

Stimuli and procedure The displays consisted of random-dot patterns that portrayed a diamond-shaped central surface surrounded by a stationary surface, as shown in Fig. 4a. Dots in the central region shifted to the left or right at a constant speed of 2 deg/s or remained stationary. When in motion, the diamond-shaped border either moved with

the dots, depicting a surface moving in front of the stationary background, or the borders remained stationary, as if the central surface were moving behind a diamond-shaped stationary aperture (the dashed border shown in Fig. 4a did not appear in the actual displays). Motion-defined shapes with moving borders are not localized as well as those with stationary borders (Regan, 1989), so the perception of depth order for the case of moving borders might also be weakened.

Stereo image pairs were created in which either disparity was the same (zero) for the entire pattern or there was a small difference in disparity between the central and surrounding regions. In the latter case, one surface had zero disparity and the other had a small positive disparity of 6'. The reliability of the stereo cue was varied by including different fractions of dots whose positions in the left and right image were not correlated. In particular, the stereo images contained 20%, 60%, or 100% dots whose positions had a coherent disparity, with the remaining dots being uncorrelated between the two images. This created an impression of dots floating in front of or behind the depth associated with the specified disparity. Red dots were displayed on a black background, and each frame contained an average of 5,600 dots, giving an average dot density of 14 dots/deg². The overall pattern subtended 20° × 20° of visual arc, and each side of the diamond subtended 7° of visual arc. When motion occlusion was present, disappearing dots were replaced so that a constant density of dots was maintained. On each trial, a white fixation cross with zero disparity appeared alone, and the subject pressed a key to display the moving dot pattern. After 0.5 s the pattern

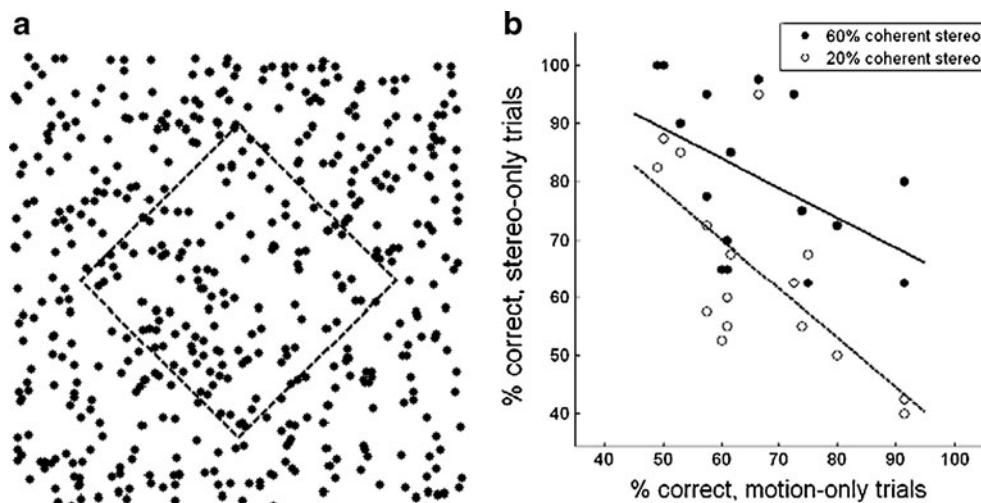


Fig. 4 (a) Display used in Experiment 3. A diamond-shaped area of dots was displayed with a different stereo disparity and direction of motion relative to the surrounding area. The dashed line did not appear in the actual display. (b) Data for single-cue conditions for Experiment 3. The percentages correct for stereo-only conditions are

plotted against the percentages correct for motion-only conditions. Individual data (mean performance across both blocks) are shown for the 60% (filled circles) and 20% (open circles) stereo coherence conditions. Regression lines are also shown for the 60% (solid line) and 20% (dashed line) stereo coherence conditions

disappeared, and the subject indicated with a second keypress whether the central diamond appeared to be in front of or behind the surrounding stationary surface.

This experiment was divided into two blocks, each containing 10 repetitions of 22 conditions. Ten conditions were common to both blocks and provided only a single cue. In six of these conditions, only the stereo cue was provided, with the central surface displayed in front or in back and with three percentages of dots with a coherent stereo disparity (20%, 60%, and 100%). Four conditions included the motion occlusion cue alone, with motion to the left or right and borders that were moving or stationary. The remaining 12 conditions in both blocks contained both stereo and motion occlusion cues. Again, the stereo cue indicated that the central surface was in front or in back and had three different percentages of dots with a coherent stereo disparity. The central surface moved to the left or the right. In the first block, the depth order indicated by the movement of the borders of the diamond-shaped region agreed with the order defined by stereo disparity. In the second block, the motion occlusion indicated the opposite depth order.

Results

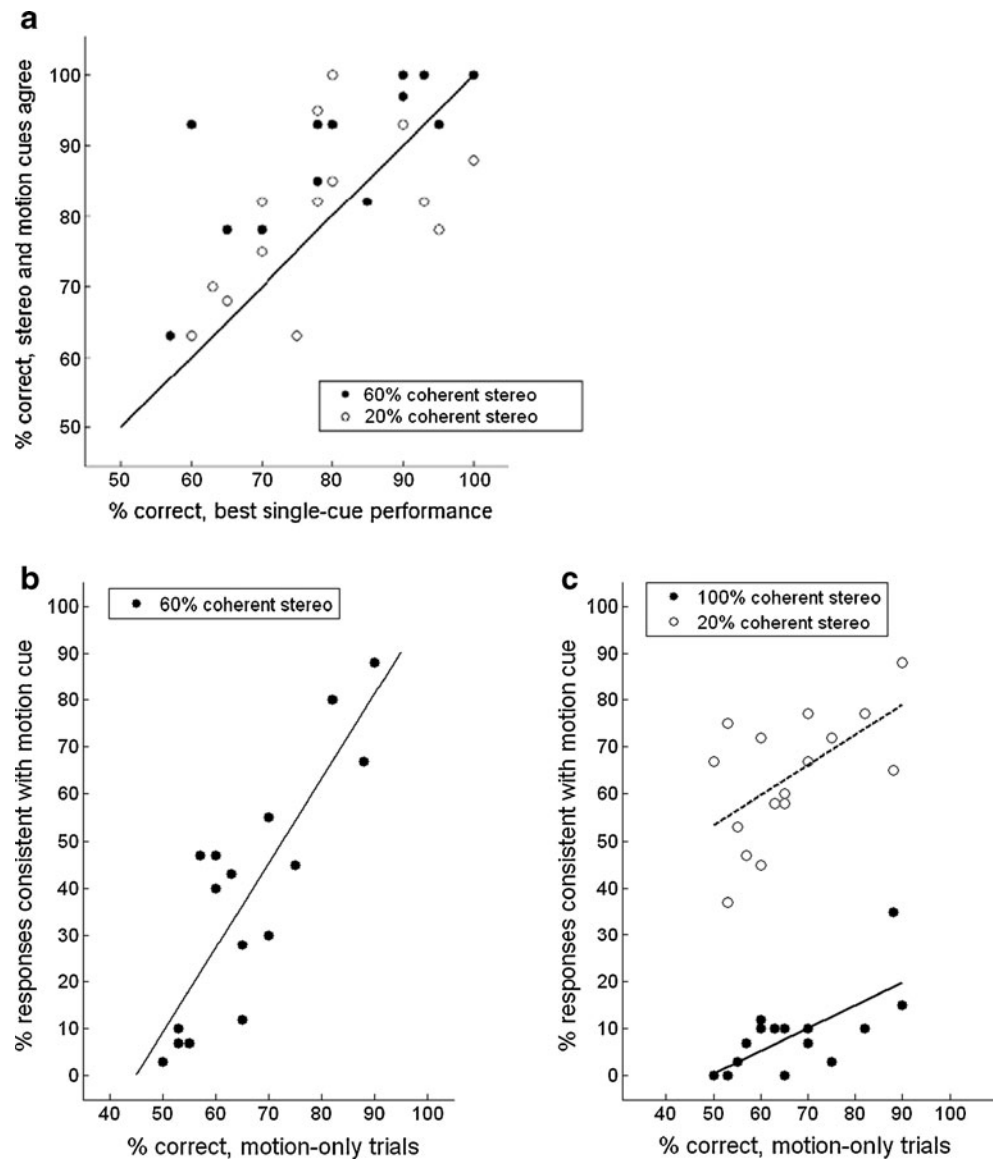
We observed large differences in performance across subjects on the single-cue conditions. For conditions in which the boundaries were defined by motion occlusion alone, the percentages of correct depth order judgments across the two blocks ranged from 49%–92% for individual subjects, with similar performance in both blocks. For stereo-only conditions, mean performance across subjects was 98% correct for the 100%-coherent condition, but for the 60%-coherent case, individual performance ranged from 63% to 100% correct. For the 20%-coherent case, individual performance ranged from 40% to 95% correct. Performance on stereo-only conditions was also similar across blocks. Subjects' performance was better for stereo-only trials on which the central surface was in back, for both the 20% and 60% stereo coherence conditions (on average, the percent correct was 12% higher with the central surface in back). Figure 4b shows the data for individual subjects, with the percentages of correct responses in the stereo-only trials plotted against the percentages correct in motion-only trials. Filled and open circles show data for the 60%- and 20%-coherent stereo conditions, respectively, and the solid and dashed lines show regression lines for these two data sets, respectively. Subjects who performed especially well on the 20%-coherent stereo condition tended to perform poorly on the motion-only conditions, and those who performed especially well on the motion-only trials performed poorly on this

challenging stereo condition. This negative correlation was significant (slope = -0.85 , $r^2 = .48$, $p = .0014$). Performance on the stereo-only trials with 60%-coherent stereo was also negatively correlated with performance on the motion-only trials, although the correlation was weaker (slope = -0.51 , $r^2 = .24$, $p = .028$).

Figure 5 shows individual data for conditions in which motion and stereo cues were combined. Figure 5a shows the extent to which performance was improved when the two cues indicated the same depth order. For each subject, we first determined which cue, stereo or motion occlusion, yielded the best performance when presented alone. In Fig. 5a, the percentages correct for trials on which stereo and motion cues indicated the same depth order are plotted against the percentages correct for the best single-cue performance. The filled and open circles show data for cue-combination trials with 60%- and 20%-coherent stereo, respectively. (There are fewer than 16 visible data points for each case due to overlap in the data for individual subjects.) If performance on trials in which the two cues were combined was determined entirely by subjects' "best" individual cue, the data would lie along the solid line shown in Fig. 5a. On average, performance in the combined-cue condition was only slightly higher than the best single-cue performance. For the 60%-coherent stereo conditions, the mean increase in performance (omitting 3 subjects whose best single-cue performance was 100% correct) was 8.6% (SD 9.4%). For the 20%-coherent stereo conditions, the mean increase in performance was 3.1% (SD 10.1%), and a few subjects showed a slight decrease in performance when consistent cues were combined in this case. This indicates only a slight enhancement in performance on this task when multiple cues reinforced one another.

Figure 5b and c show performance for individual subjects in the cue-conflict conditions. The data reveal large differences between subjects in the percentages of depth order judgments consistent with the motion occlusion cue versus the stereo cue. For the 60%-coherent stereo conditions, the percentages of depth order judgments consistent with the motion occlusion cue ranged from 2% to 88%. In this case, there was a strong correlation between the percentage of judgments consistent with the motion occlusion cue and performance with the motion cue alone, as shown in Fig. 5b (slope = 1.8 , $r^2 = .72$, $p < .001$). Similar data for the 100%- and 20%-coherent stereo conditions are shown as filled and open circles, respectively, in Fig. 5c. All subjects were much more likely to select the depth order specified by stereo when the stereo cue was 100% coherent, and more likely to select the depth order specified by motion occlusion when the stereo cue was very weak (only 20% coherent), so there was less variation in performance across subjects for these two conditions.

Fig. 5 Individual data for the cue-combination conditions of Experiment 3. (a) The percentages of correct depth-order judgments for conditions in which stereo and motion cues were combined and specified the same depth order are plotted against the best performance achieved for a single cue alone. Filled and open circles show data for the 60%- and 20%-coherent stereo conditions, respectively. The solid line shows the expected data if performance on the cue-combination conditions were determined entirely by the best individual cue. (b and c) Individual data for conditions in which stereo and motion cues indicated conflicting depth orders. The percentages of judgments consistent with the motion cue are plotted against performance in the motion-only conditions. (b) Data for the 60% stereo coherence condition, with regression line. (c) Data for the 100% (filled circles) and 20% (open circles) stereo coherence conditions, with regression lines



Discussion

Individual subjects differed greatly in their ability to judge relative depth at object boundaries from stereo and motion occlusion cues in isolation. These differences were revealed when the task became very difficult—for example, when the stereo cue was weakened by adding a large fraction of uncorrelated dots. The motion occlusion cue in our displays might have been weakened by the short presentation time, small change in motion across the object borders, or peripheral placement of the borders. Most subjects exhibited relatively poor performance on both the motion-only and the challenging stereo-only (20% coherence) tasks, in the range of only 50% – 75% correct. We found evidence for a negative correlation between performance on these two tasks; subjects who had the best performance with one cue tended to have the worst performance with the

other. These differences had consequences for the interpretation of displays with conflicting stereo and motion cues. As the stereo cue was progressively weakened from 100% to 20% coherent, subjects who performed better on the stereo-only task, subjects who performed better on the stereo-only task continued to select the depth order specified by stereo when a conflicting motion cue was added to the display. In contrast, those who performed better on the motion-only task appeared to quickly abandon the stereo cue in favor of the motion occlusion cue. These subjects had also tended to select the depth order consistent with motion occlusion when a conflicting speed cue was added in Experiment 1. The data for the 60%-coherent stereo condition (Fig. 5b) highlight the range of behavior observed in the present experiment. When the stereo and motion cues indicated the same depth order, for most subjects performance was only slightly enhanced relative to the subjects' best performance with the cues in isolation.

For a few subjects, the combination of two consistent cues led to a large increase in performance, similar to the enhancement observed for the recovery of 3-D shape from multiple cues (e.g., Bühlhoff & Mallot, 1988) and depth volume from stereo and motion cues (e.g., van Ee & Anderson, 2001).

Experiment 4

In the previous experiment, only the reliability of the stereo cue was varied. In this experiment, the reliability of both the stereo and motion cues was varied. The experiment used displays similar to those in Experiments 1 and 2, in which the random-dot surfaces moved toward the observer and included a relative speed cue to depth order. This experiment also examined whether the relative influences of individual cues to depth order depend on a broader context than the trial-by-trial variation in reliability. In particular, the experiment was divided into three blocks, in which (1) only the stereo cue was degraded, (2) only the motion cue was degraded, and (3) the reliability of both cues was varied. If the weight given to a particular cue depends on context, we might expect that, overall, there would be less reliance on the stereo cue in the session in which only the stereo cue was degraded, and similarly, less reliance on the motion cue when only this cue was degraded.

Method

Subjects A group of 17 naïve subjects with normal or corrected-to-normal vision participated in this experiment. All of these subjects had participated in Experiments 1 and 2, and 16 of them had participated in Experiment 3.

Stimuli and procedure As noted above, this experiment was divided into three blocks. In the first two blocks, only one of the two cues was degraded, while the other remained strong throughout the block. In the third block, conditions with weaker stereo or motion cues were intermixed. The three blocks contained overlapping conditions in which both cues were strong. Similar to those in Experiments 1 and 2, the displays portrayed two frontoparallel surfaces placed side by side and moving toward the observer, with a stationary vertical border at the center of the image. One of the surfaces had an additional horizontal movement toward the central border, resulting in motion occlusion. There was a small difference in stereo disparity between the two surfaces, and disparity remained constant as the surfaces moved forward.

In the first block, the motion occlusion cue was weakened by reducing the difference in the 3-D direction

of motion between the two surfaces. This was accomplished by having one surface move in depth only, and varying the horizontal component of translation of the second surface toward the center of the image. As this horizontal shift was reduced, fewer dots disappeared along the central border and the size of the change in image velocity across the border was reduced, weakening the impression of a depth change due to motion occlusion. This manipulation maintained the coherent flow indicating motion of the surfaces in depth, while targeting the strength of the percept of a depth change at the border. This block contained 10 repetitions of 16 conditions. One surface had a small positive disparity of 6', while the other had a small negative disparity of -6', and the front surface could appear on the left or the right. Disparity remained constant over time. The initial TTC was 1.5 s for both surfaces. The 3-D direction of motion of the back, occluded surface was either 1.4°, 2.8°, 4.2°, or 5.6° to the left or right of the viewer. The occluded surface could appear on the left or the right, giving eight combinations of motion parameters. In half of the trials, the motion occlusion and stereo disparity cues indicated a consistent depth order, while in the other half, the two cues were in conflict.

In the second block, the stereo cue was weakened by varying the fractions of dots in the left and right images whose positions were not correlated across views, similar to the previous experiment. There were 10 repetitions of 16 conditions. Both surfaces had an initial TTC of 1.5 s, with one moving in depth only and the other moving in a 3-D direction of motion that was 5.6° to the left or right of the viewer. The stereo disparities of the surfaces were 6' or -6', with the front surface on the left or right, but the percentage of dots with this disparity was varied over four values: 25%, 50%, 75%, and 100%. Trials were divided evenly between those providing consistent and conflicting information about depth order from stereo and motion cues.

The final block contained 20 conditions. The stereo disparity of each surface was 6' or -6', with the front surface on the left or right. The occluding surface, which moved in depth only, could also appear on the left or right, and the surfaces had the same initial TTC of 1.5 s. There were five combinations of relative strengths of the stereo and motion cues: The pairs of values for the difference in 3-D direction of motion and percentage of dots with a coherent disparity were (5.6°, 100%), (4.2°, 100%), (2.8°, 100%), (5.6°, 75%), and (5.6°, 50%).

In all trials of this experiment, the displays contained an average of 2,000 dots per frame and subtended 20° × 20° of visual arc, giving an average dot density of 5 dots/deg². A fixation cross appeared initially, and following a keypress the images immediately appeared in motion for 0.5 s and disappeared. Subjects indicated whether the left or right surface appeared in front with a second keypress.

Results

For conditions in which stereo and motion cues indicated the same depth order, subjects made the correct judgments in 95% – 100% of the trials. For conditions in which these cues indicated opposite depth orders, we again observed differences between subjects in the resolution of this conflict. Overall, the percentage of judgments consistent with a particular cue decreased as the cue became more unreliable, but only the effect of stereo degradation was significant. This is shown in Fig. 6a. The curve with filled circles shows the mean data for all subjects for the cue-conflict conditions in the second block, in which the stereo cue was degraded. The percentages of depth order judgments consistent with the stereo cue are plotted against the percentages of dots with the coherent stereo disparity (horizontal axis at the bottom of the figure). A one-way ANOVA showed a significant effect of stereo coherence on depth order judgments in this case [$F(3, 64) = 5.35, p = .002$]. The open circles show mean data for all subjects for the cue-conflict conditions in the first block, in which the motion cue was degraded. The difference in 3-D direction of motion between the two surfaces is shown along the horizontal axis at the top of the figure. In this case, the mean data do not show a significant effect of motion degradation [$F(3, 64) = 1.09, p = .38$]. The condition with 100%-coherent stereo and a 5.6° difference in 3-D direction of motion was common to both blocks. The data in Fig. 6a

show that in the block in which only the motion cue was degraded, the percentage of judgments consistent with the stereo cue for this common condition was 20% higher, on average, relative to the percentage of judgments consistent with stereo in the block in which only the stereo cue was degraded.

Individual data that convey some of the range of behavior observed in the cue-conflict conditions are shown in Fig. 6b and c. Figure 6b shows individual data for the first block, in which only motion was degraded, and Figure 6c shows data for the same subjects for the second block, in which only stereo was degraded. In the case of Subject 3 (diamonds), depth order judgments were dominated by the stereo cue and showed little or no variation with degradation of either cue. Subject 4 (triangles) exhibited the opposite behavior, in which judgments were dominated by the motion occlusion cue, and also showed no variation in performance with degradation of either cue. These 2 subjects were among the groups in Experiment 3 who exhibited highly asymmetric performance on the stereo-only and motion-only conditions. For Subjects 1 and 2 (circles and squares) in Fig. 6b and c, performance was closer to chance when the two cues were strong, but when one of the cues was very weak, judgments favored the stronger cue. The performance of Subject 1 on the motion-only and the most challenging stereo-only conditions of Experiment 3 had been less than 75% correct. In general, performance for individual subjects was similar on

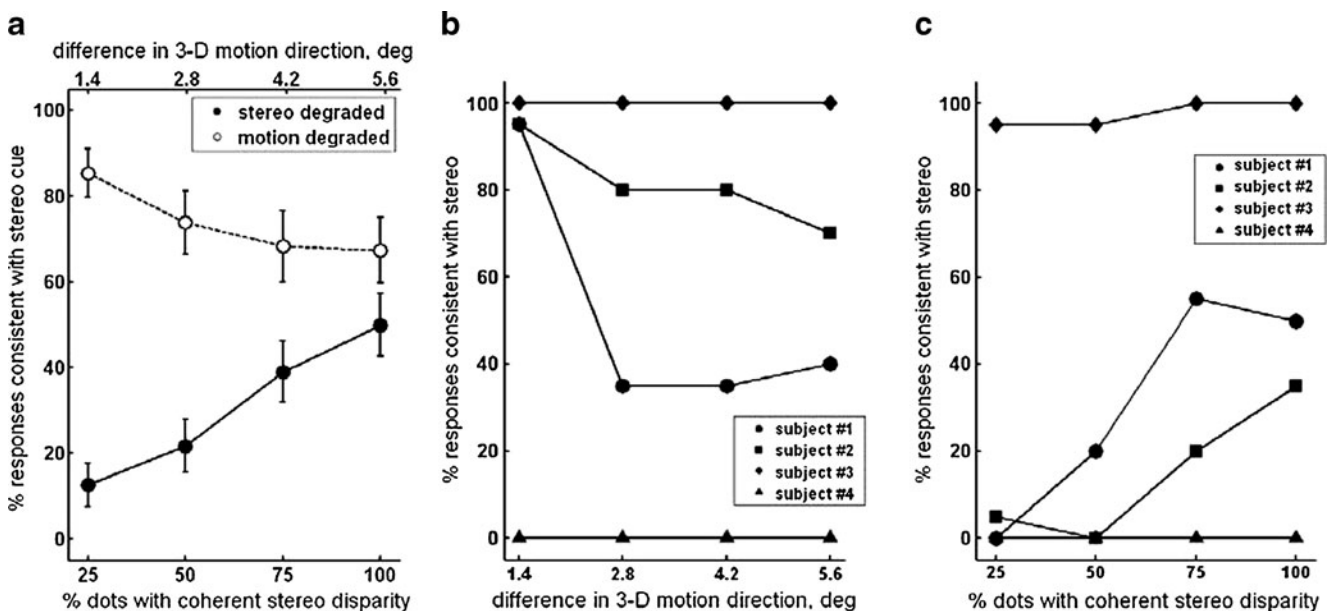


Fig. 6 Data from Experiment 4. (a) Mean data for cue-conflict conditions for all subjects are shown for the first block (open circles, dashed line), in which the motion cue was degraded. The percentages of trials in which depth order judgments were consistent with the stereo cue are plotted against the difference in 3-D direction of motion of the two surfaces (top horizontal axis). Mean data for all subjects are

also shown for the second block (filled circles, solid line), in which the stereo cue was degraded. Here, the percentages of judgments consistent with the stereo cue are plotted against the percentages of dots with coherent disparity (bottom horizontal axis). Error bars indicate standard errors. (b and c) Individual data are shown for 4 subjects for the first block (b) and the second block (c)

the common conditions of Blocks 1 and 3 and the common conditions of Blocks 2 and 3.

Discussion

The results of this experiment reinforce the observations from [Experiment 3](#), that large individual differences exist in the way subjects resolve conflicts between stereo and motion occlusion cues to depth order. All of the subjects in this study were capable of using either cue to infer depth order, but they exhibited extremes of behavior in cue-conflict conditions, from dominance of the stereo cue to dominance of the motion occlusion cue. For some subjects, this dominance persisted even when the dominant cue was severely degraded. Many subjects showed only a slight preference for one cue or the other when conflicting cues were combined, and these subjects also tended to vary in performance when one or the other cue was degraded. We found weak evidence of a contextual influence. Performance in a common condition in the first two blocks of the experiment showed stronger preference for the stereo cue when only the motion cue was degraded, and greater preference for the motion cue when only the stereo cue was degraded within the block. There was little change in behavior, however, between blocks with only one cue degraded and the third block, in which both cues were degraded in different trials.

General discussion

We have presented the results of experiments that explored the interaction between constraints on depth order provided by relative image speed, motion occlusion, and stereo disparity. [Experiment 1](#) confirmed the expectation that in the absence of explicit cues to depth order from motion occlusion or stereo disparity, a surface with faster image velocities appears closer in depth. Informal reports from a few subjects suggested that this percept may have emerged over time, in part as a consequence of the interpretation of faster image velocities as resulting from faster speeds of motion in depth. The model for the recovery of 3-D structure from motion proposed by Hildreth et al. (1995; Ullman, 1984) suggests one way that this percept might emerge.

[Experiment 1](#) also showed that for most subjects, motion occlusion, conveyed by the disappearance of image features along the border of an occluding surface, results in a perceived depth order that overrides the depth order suggested by the relative image speeds of features within adjacent surfaces. This behavior is consistent with the observations of Braunstein et al. (1982) and Ono et al. (1988) regarding perceived depth order in displays of 3-D

object rotation and displays created from lateral movements between the viewer and surfaces. For most subjects, the depth order implied by motion occlusion can effectively “veto” that obtained from the interpretation of relative image speed. In Hildreth et al.’s (1995) model, the surface reconstruction process can incorporate such depth order constraints from motion occlusion. The results of [Experiment 1](#) also showed that depth order judgments for some subjects are more strongly influenced by the relative speed cue, in some cases ignoring the motion occlusion cue. Even when the perception of motion occlusion was strong, this cue alone tended to elicit the sensation of only a small depth difference between the two surfaces.

As a result of the interaction between motion occlusion and relative image speed, a surface with faster image speeds can be perceived as moving faster in depth, while its position in depth is constrained by a slower-moving surface if the motion occlusion cue indicates that the slower surface is in front. In this case, the faster-moving surface in back appears to continually move closer to the slower, occluding surface, without ever passing the front surface. This appearance suggests some perceptual decoupling between the representation of motion in depth and position in depth (Rokers, Cormack, & Huk, 2008); a surface can appear to move forward continually without advancing its position in depth. Engel, Remus, and Sainath (2006, in “Movie 4”) provided a perceptual demonstration also showing that motion occlusion can alter an object’s perceived 3-D motion trajectory. In their display, a laterally moving sphere of constant size was perceived as moving in depth as a consequence of occlusion.

The results of [Experiment 2](#) show that depth from stereo disparity can also impose a strict constraint on depth order. In our displays, stereo disparity remained constant over time, and the simultaneous appearance of an expanding pattern of image motion resulted in a strong percept of movement in depth. A surface with faster image speeds is perceived as moving faster in depth, but it appears that for most subjects the constraint on depth order indicated by stereo disparity cannot easily be overridden. A faster-moving surface whose disparity places it behind a slower-moving surface will appear to approach the slower front surface continually, without passing it in depth. This again suggests a separation of the representations of depth and movement in depth.

Edwards and Badcock (2003) also examined the interaction of stereo and motion cues, with the aim of testing whether motion alters the percept of a static stereo disparity signal. They measured the perceived depth of random-dot surfaces with constant stereo disparity, in which the dots also moved in an expanding or contracting pattern. The speeds of movement were constant over the surface, in order to weaken the percept of movement in depth of the

surfaces. The added expanding motion caused the perceived depth of a surface to shift closer to the viewer, while added contracting motion shifted perceived depth away from the viewer. In the displays for our experiments, in which image speeds varied across the surfaces as expected for movement in depth, the surfaces elicited a strong impression of movement in depth. The surfaces did not appear to remain stationary in space, as indicated by the stereo cue. It is likely that the biases in depth location observed in the experiments of Edwards and Badcock were related to the movements in depth observed in our displays.

Experiments 3 and 4 examined the direct interaction between motion occlusion and stereo disparity cues to depth order. In principle, two cues providing weak but consistent information about the relative depths of two surfaces could reinforce or enhance one another, resulting in a more consistent depth percept relative to that obtained from each cue in isolation. Rivest and Cavanagh (1996), for example, showed that the assessment of boundary location is enhanced by the combination of multiple color, motion, and texture cues. The results of Experiment 3 showed that for most subjects, there was only a small increase in performance when stereo and motion occlusion cues indicated the same depth order for the surfaces, relative to the level of performance that subjects were able to reach with a single cue.

We observed substantial variation in performance when the stereo and motion occlusion cues were placed in conflict. This range of performance was also reflected in subjects' abilities to judge depth order from single cues. A few of our subjects performed extremely well with a severely degraded stereo stimulus but performed poorly in displays that contained only a weak motion occlusion cue. These subjects also strongly favored the stereo cue when the two cues were placed in conflict. In contrast, a few subjects performed extremely well on displays that contained only the motion occlusion cue and performed poorly on severely degraded stereo displays. These subjects strongly favored the motion cue in cue-conflict conditions.

Individual differences in stereo and motion performance have been observed previously, although few studies have compared performance across these two cues. The data presented in studies by Møller and Hurlbert (1996) and Smith and Curran (2000) for motion segmentation tasks, for example, suggested large differences in the speed thresholds (Møller & Hurlbert, 1996) and direction thresholds (Smith & Curran, 2000) required for the reliable detection of motion boundaries in briefly viewed dynamic random-dot patterns. In a study of over 60 subjects, Nefs, O'Hare, and Harris (2010) found differences in subjects' ability to judge motion-in-depth from the two cues of changing disparity over time and interocular velocity differences. They provided some evidence that subjects who are

especially good at using one of these cues tended to be poor at using the other. Van Ee and Anderson (2001) observed enhanced performance in the judgment of depth volume in dynamic random-line stereograms when stereo and motion cues were combined. Van Ee (2003) later examined whether stereoanomalous viewers, with poor stereo discrimination ability in the near or far disparity regions, exhibited this same enhancement when stereo and motion were combined in the disparity range that the viewers could process effectively. He found, however, that stereoanomalous viewers tend to rely completely on motion for depth judgments, rather than taking advantage of their limited stereo capability as well. The subjects in our experiments who were especially good at judging depth order from motion occlusion, and poor at making similar judgments in the challenging stereo conditions, also tended to abandon the stereo cue when the motion cue was placed in conflict with a weaker stereo cue. It is possible that these subjects were also stereoanomalous—we were not able to determine this from our experiments.

Models of cue integration that compute a weighted combination of depth values obtained from multiple cues (e.g., Bruno & Cutting, 1988; Landy et al., 1995; Yuille & Bülthoff, 1996) adjust the weights associated with individual cues depending on the reliability, ambiguity, or consistency of the cues (Jacobs, 2002). Reliability is typically defined in terms of the variance of the inferences about depth that can be derived from a particular cue. Most studies of depth-cue combination have used judgments of absolute depth or quantitative aspects of depth differences. In our experiments, subjects were simply asked to judge depth order or the sign of the temporal change in the depth difference between the two surfaces. Depth information from stereo and motion could initially be combined within the two surface regions in a way that incorporated the variance of stereo disparity and image motion, prior to making an assessment of depth order or change in relative depth. In this way, performance on qualitative depth judgments could still depend on a weighted combination of depth information from multiple cues. Alternatively, ordinal depth information from motion cues could be transformed, prior to cue combination, to a quantitative representation of relative or absolute depth that is more commensurate with that provided by stereo disparity (Landy et al., 1995). The information from the two cues could then be combined using a weighted adjustment, as described above.

Conclusion

Subjects are able to use the individual cues of speed, occlusion, and stereo disparity to infer depth order and

change in the relative depths of two surfaces. When these cues are in conflict, we find significant individual differences in the way this conflict is resolved. For some subjects, one cue dominates their perception. In the case of conflicting motion occlusion and stereo cues, the emergence of a dominant cue appears to be rooted in a large discrepancy between subjects' abilities to process motion and stereo cues in isolation. The perception of relative depth appears to favor the 3-D structure conveyed by a more reliable cue when one cue is degraded. For most subjects, a weak relative depth cue can be enhanced by the presence of another weak cue that reinforces the same depth relationship. Further study is needed to explore the underlying basis for individual differences in depth-cue combination.

Author Note We thank Jeremy Wilmer for helpful feedback on a draft of the manuscript. This work was supported by a grant from the National Science Foundation to Constance Royden (NSF Grant #IOS-0818286).

References

- Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review*, *101*, 414–445. doi:10.1037/0033-295X.101.3.414
- Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Reviews of Neuroscience*, *28*, 157–189.
- Bradley, D. C., & Andersen, R. A. (1998). Center-surround antagonism based on disparity in primate area MT. *Journal of Neuroscience*, *18*, 7552–7565.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics*, *31*, 261–267. doi:10.3758/BF03202532
- Bredfeldt, C. E., & Cumming, B. G. (2006). A simple account of cyclopean edge responses in Macaque V2. *Journal of Neuroscience*, *26*, 7581–7596.
- Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, *117*, 161–170. doi:10.1037/0096-3445.117.2.161
- Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America, A*, *5*, 1749–1758.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). New York: Academic Press.
- Edwards, M., & Badcock, D. R. (2003). Motion distorts perceived depth. *Vision Research*, *43*, 1799–1804.
- Egnal, G., & Wildes, R. P. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1127–1133.
- Engel, S. A., Remus, D. A., & Sainath, R. (2006). Motion from occlusion. *Journal of Vision*, *6*(5), 9:649–652. doi:10.1167/6.5.9
- Feldman, D., & Weinshall, D. (2008). Motion segmentation and depth ordering using an occlusion detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*, 1171–1185.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J., Kaplan, G. A., Reynolds, H. N., & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception & Psychophysics*, *5*, 113–116.
- Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America, A*, *4*, 503–518.
- Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research*, *49*, 2666–2685. doi:10.1016/j.visres.2009.06.021
- Hildreth, E. C., Ando, H., Andersen, R. A., & Treue, S. (1995). Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research*, *35*, 117–137.
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth: Vol 2. Depth perception*. Toronto: Porteous.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–350. doi:10.1016/S1364-6613(02)01948-4
- Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics*, *6*, 193–198.
- 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. doi:10.1016/0042-6989(94)00176-M
- Liu, Y., Vögels, R., & Orban, G. A. (2004). Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *Journal of Neuroscience*, *24*, 3795–3800.
- Longuet-Higgins, H. C., & Prazdny, K. (1981). The interpretation of moving retinal images. *Proceedings of the Royal Society B*, *208*, 385–397.
- Marcar, V. L., Raigel, S. E., Xiao, D., & Orban, G. A. (2000). Processing of kinetically defined boundaries in areas V1 and V2 of the Macaque monkey. *Journal of Neurophysiology*, *84*, 2786–2798.
- Marcar, V. L., Xiao, D.-K., Raigel, S. E., Maes, H., & Orban, G. A. (1995). Processing of kinetically defined boundaries in the cortical motion area MT of the Macaque monkey. *Journal of Neurophysiology*, *74*, 1258–1270.
- Møller, P., & Hurlbert, A. C. (1996). Psychophysical evidence for fast region-based segmentation processes in motion and color. *Proceedings of the National Academy of Sciences*, *93*, 7421–7426.
- Mysore, S. G., Vögels, R., Raigel, S. E., & Orban, G. A. (2006). Processing of kinetic boundaries in Macaque V4. *Journal of Neurophysiology*, *95*, 1864–1880.
- Nefs, H., O'Hare, L., & Harris, J. (2010). Two independent mechanisms for motion-in-depth perception: evidence from individual differences. *Frontiers in Psychology*, *1*, 155. doi:10.3389/fpsyg.2010.00155.
- Ono, H., Rogers, B. J., Ohmi, M., & Ono, M. E. (1988). Dynamic occlusion and motion parallax in depth perception. *Perception*, *17*, 255–266.
- Orban, G. A. (2008). Higher order processing in macaque extrastriate cortex. *Physiological Reviews*, *88*, 59–89.
- Patterson, R., Cayko, R., Short, G. L., Flanagan, R., Moe, L., Taylor, E., et al. (1995). Temporal integration differences between crossed and uncrossed stereoscopic mechanisms. *Perception & Psychophysics*, *57*, 891–897.
- Pong, T. C., Kenner, M. A., & Otis, J. (1990). Stereo and motion cues in preattentive vision processing—Some experiments with

- random-dot stereographic image sequences. *Perception*, *19*, 161–170.
- Proffitt, D. R., Bertenthal, B. I., & Roberts, R. J. (1984). The role of occlusion in reducing multistability in moving point-light displays. *Perception & Psychophysics*, *36*, 315–323.
- Qiu, F., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron*, *47*, 155–166.
- Regan, D. (1989). Orientation discrimination for objects defined by relative motion and objects defined by luminance contrast. *Vision Research*, *29*, 1389–1400. doi:10.1016/0042-6989(89)90194-6
- Rivest, J., & Cavanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research*, *36*, 53–66. doi:10.1016/0042-6989(95)00056-6
- Rokers, B., Cormack, L. K., & Huk, A. C. (2008). Strong percepts of motion through depth without strong percepts of position in depth. *Journal of Vision*, *8*(4), 6:1–10. doi:10.1167/8.4.6
- Royden, C. S., Baker, J. F., & Allman, J. (1988). Perceptions of depth elicited by occluded and shearing motions of random dots. *Perception*, *17*, 289–296.
- Smith, A. T., & Curran, W. (2000). Continuity-based and discontinuity-based segmentation in transparent and spatially segregated global motion. *Vision Research*, *40*, 1115–1123.
- Tadin, O., & Lappin, J. S. (2005). Linking psychophysics and physiology of center-surround interactions in visual motion processing. In M. R. M. Jenkin & L. R. Harris (Eds.), *Seeing spatial form* (pp. 279–314). Oxford: Oxford University Press.
- Thompson, W. B., Mutch, K. M., & Berzins, V. A. (1985). Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-7*, 374–383.
- Tyler, C. W. (2004). Binocular vision. In B. E. Rogowitz, T. N. Pappas, & S. J. Daly (Eds.), *Duane's foundations of clinical ophthalmology*. Philadelphia: Lippincott.
- Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. *Perception*, *13*, 255–274.
- van den Berg, A. V., & Beintema, J. A. (2000). The mechanism of interaction between visual flow and eye velocity signals for heading perception. *Neuron*, *26*, 747–752.
- van Ee, R. (2003). Correlation between stereoanomaly and perceived depth when disparity and motion interact in binocular matching. *Perception*, *32*, 67–84. doi:10.1068/p3459
- van Ee, R., & Anderson, B. L. (2001). Motion direction, speed and orientation in binocular matching. *Nature*, *410*, 690–694. doi:10.1038/35070569
- von der Heydt, R., Qiu, F. T., & He, Z. J. (2003). Neural mechanisms in border ownership assignment: Motion parallax and Gestalt cues [Abstract]. *Journal of Vision*, *3*(9), 666a.
- von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, *40*, 1955–1967.
- Wilcox, L. M., & Allison, R. S. (2009). Coarse-fine dichotomies in human stereopsis. *Vision Research*, *49*, 2653–2665. doi:10.1016/j.visres.2009.06.004
- Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, *21*, 561–579.
- Wismeijer, D. A., Erkelens, C. J., van Ee, R., & Wexler, M. (2010). Depth cue combination in spontaneous eye movements. *Journal of Vision*, *10*(6), 25:1–15. doi:10.1167/10.6.25
- Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics*, *41*, 53–59. doi:10.3758/BF03208213
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge: Cambridge University Press.
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*, 6594–6611.
- Ziegler, L. R., & Hess, R. F. (1997). Depth perception during diplopia is direct. *Perception*, *26*, 1225–1230.