Da Vinci decoded: Does da Vinci stereopsis rely on disparity?

Inna Tsirlin
Centre for Vision Research, Department of Psychology, York University, Toronto, Ontario, Canada

Laurie M. Wilcox
Centre for Vision Research, Department of Psychology, York University, Toronto, Ontario, Canada

Robert S. Allison
Centre for Vision Research, Department of Psychology and Department of Computer Science & Engineering, York University, Toronto, Ontario, Canada

In conventional stereopsis, the depth between two objects is computed based on the retinal disparity in the position of matching points in the two eyes. When an object is occluded by another object in the scene, so that it is visible only in one eye, its retinal disparity cannot be computed. Nakayama and Shimojo (1990) found that a percept of quantitative depth between the two objects could still be established for such stimuli and proposed that this percept is based on the constraints imposed by occlusion geometry. They named this and other occlusion-based depth phenomena “da Vinci stereopsis.” Subsequent research found quantitative depth based on occlusion geometry in several other classes of stimuli grouped under the term da Vinci stereopsis. However, Nakayama and Shimojo’s findings were later brought into question by Gillam, Cook, and Blackburn (2003), who suggested that quantitative depth in their stimuli was perceived based on conventional disparity. In order to understand whether da Vinci stereopsis relies on one type of mechanism or whether its function is stimulus dependent we examine the nature and source of depth in the class of stimuli used by Nakayama and Shimojo (1990). We use three different psychophysical and computational methods to show that the most likely source for depth in these stimuli is occlusion geometry. Based on these experiments and previous data we discuss the potential mechanisms responsible for processing depth from monocular features in da Vinci stereopsis.

Keywords: stereopsis, da Vinci stereopsis, monocular occlusions, depth perception, quantitative depth, mechanisms, cross-correlation


Introduction

Leonardo da Vinci was one of the first scientists to discuss the phenomenon of monocular occlusions—areas visible only to one eye during binocular viewing (da Vinci, trans. 1877; Wade, Ono, & Lillakas, 2001). These areas arise due to physical occlusion of objects by other objects in the world in one eye’s image. Because there is no match in the fellow eye, monocularly occluded features might be thought of as a challenge to solving the correspondence problem and therefore a hindrance to stereopsis; however, a growing body of research suggests that they are an integral part of binocular depth perception (for review, see Harris & Wilcox, 2009). It has even been proposed that the development of forward-facing eyes in animals was governed by the advantage provided by the ability to see monocularly occluded areas in cluttered environments (Changizi & Shimojo, 2008).

Some of the early work in this field showed that monocular areas can induce the perception of depth in the absence of conventional disparity (Kaufman, 1965), give rise to illusory occluding contours (Lawson & Gulick, 1967), and increase the processing speed of depth from disparity (Gillam & Borsting, 1988).

In 1990, Nakayama and Shimojo demonstrated that monocular occlusions give rise to quantitative as well as qualitative depth percepts. Quantitative depth in this case refers to a monotonic increase in the magnitude of perceived depth as the size of the monocularly occluded region is increased. In their stimulus, shown in Figure 1A, a binocular rectangle was presented with zero disparity. In one of the half-images a thin monocular bar was positioned next to the rectangle. When the bar was positioned to the right of the rectangle in the right eye, the configuration was consistent with it being located behind the rectangle in depth and occluded by the rectangle in the left eye. When the bar was positioned to the right of the rectangle in the left eye,
the configuration was consistent with it being located in front of the rectangle but camouflaged against the rectangle in the right eye. In both cases, as shown in Figure 1B, the location in depth of the monocular bar is partially constrained by the line of sight from the eye that does not see the bar (solid black line). The minimum possible depth of the bar is determined by the distance between the outermost edges of the bar and the rectangle. Note that the maximum possible depth is not constrained.

Nakayama and Shimojo (1990) suggested that the visual system might use this minimum depth constraint to localize the monocular bar in depth. Accordingly, they proposed that the perceived depth of the monocular bar should increase as occlusion width increases. This was found to be true in occlusion arrangements (but not in camouflage arrangements). In the same article they showed that monocular occlusions can induce the perception of illusory occluding contours, and earlier, the same authors demonstrated that monocular occlusions play a role in rivalry (Shimojo & Nakayama, 1990) and can induce percep-

tion of depth in moving stimuli (Shimojo, Silverman, & Nakayama, 1988). In a nod to Leonardo da Vinci’s observations, and to distinguish it from conventional stereopsis, Nakayama and Shimojo (1990) grouped all these monocular occlusion phenomena under the name “da Vinci stereopsis” (p. 1822). In other words, they applied the term da Vinci stereopsis to any phenomena in which depth is perceived on the basis of monocular occlusions with no involvement of disparity. For consistency, we will assume their definition of da Vinci stereopsis throughout this article.

Gillam et al. (2003) proposed that quantitative depth in the stimuli used by Nakayama and Shimojo (1990) is simply an instance of double-matching, in which the monocular bar is matched to the edge of the occluding rectangle in the other eye (which is also simultaneously matched to itself). The disparity of the double-match is therefore precisely defined and equal to the distance between the bar and the rectangle. Thus double-matching in these stimuli predicts the same perceived depth as the occlusion width, assuming the outer, matching polarity edges of the bar and rectangle are matched. Gillam et al. (2003) tested this hypothesis by comparing perceived depth from two types of stimuli: (a) a stimulus with a rectangular textured binocular occluder and a thin monocular line and (b) the same stimulus but with the line replaced by a small monocular disc (see Figure 2). They argued that the disc was unlikely to be matched to the edge of the rectangular occluder because of the differences in size and shape. They found that the monocular line stimulus produced quantitative depth percepts, while the monocular disc stimulus did not. However, Gillam et al. (2003) used a different range of occlusion widths for the disc and the line stimuli, as discussed in Experiment 1, which could have affected their results.

Quantitative depth based solely on occlusion geometry has been demonstrated in other types of da Vinci stereopsis arrangements such as illusory occluding surfaces (Anderson, 1994; Gillam & Nakayama, 1999; Grove & Gillam, 2007; Pianta & Gillam, 2003; Sachtler & Gillam, 2007; Tsirlin, Wilcox, & Allison, 2010) and moving stimuli (Brooks & Gillam, 2006; Shimojo et al., 1988). Thus, the findings of Gillam et al. (2003) suggest that da Vinci stereopsis is a stimulus-dependent phenomenon since it relies on different mechanisms for the computation of quantitative depth in different occlusion configurations. However, parsimony favors the explanation that there is a single mechanism responsible for quantitative depth in all types of geometrically constrained occlusion stimuli.

To understand whether da Vinci stereopsis relies on one type of mechanism or whether its function is stimulus dependent, we examined the nature and source of depth in the class of stimuli used by Nakayama and Shimojo (1990). We refer to this class of stimuli herein.
as “two-object arrangements” to emphasize that the occluded object is completely spatially segregated from the occluder. In Experiment 1 we repeated the studies by Nakayama and Shimojo (1990) and Gillam et al. (2003) and carefully controlled all stimulus parameters to ensure that the conditions were comparable. In Experiment 2 we probed the role of disparity in two-object arrangements, using a disparity bias technique proposed by Tsirlin, Wilcox, and Allison (2011) as a litmus test for differentiation of depth from monocular occlusions and depth from disparity. Finally, we examined the response of model disparity detectors to depth in two-object stereograms. Taken together, the results of these experiments and simulations suggest that double-matching is an unlikely explanation for depth percepts in two-object arrangements. We discuss these data and the previous literature with respect to potential mechanisms responsible for assigning depth to monocular occlusions in da Vinci stereopsis.

Experiment 1

In their experiments, Gillam et al. (2003) measured occlusion width as the distance from the edge of the occluder to the center of the monocular object instead of the outer edge of the monocular object (personal communication). Since the monocular line of Gillam et al. (2003) was only 10 arcsec wide, the range of occlusion widths when the line was the target was close to the widths specified in their article, 5 to 35 arcmin. However, for the disc that was 7 arcmin in diameter, the actual range of occlusion widths was 8.5 to 38.5 arcmin. This difference in test conditions is important since it has been shown that da Vinci stereopsis only occurs over a narrow range of occlusion widths (Gillam et al., 2003; Hakkinen & Nyman, 1996; Nakayama & Shimojo, 1990). In particular, Gillam et al. (2003) found that even in their monocular line stimulus (Figure 2B), quantitative depth was observed only for occlusion widths smaller than 12–15 arc min. Therefore, it is possible that most of the occlusion widths they used with the disc stimulus were out of the range of da Vinci stereopsis and thus failed to produce quantitative depth. Although the effective range for quantitative depth Nakayama and Shimojo (1990) found was about 30–40 arcmin for their configuration, their stimuli were almost twice the size of those used by Gillam et al. (2003). It is likely that the range of quantitative depth from monocular occlusions is affected by the size of the stimuli similarly to depth from disparity (for review, see Wilcox & Allison, 2009). Given this, it’s important to ensure that comparisons of depth from monocular occlusions are made between same-scale arrangements.

In this experiment we replicated several conditions from Nakayama and Shimojo (1990) and Gillam et al. (2003), while carefully controlling occlusion widths to equate minimum predicted depth from occlusion for all monocular stimuli.

Methods

Observers

Six observers participated in this experiment. Two were authors (IT and LW) and others were naïve as to the purpose of the experiment. All observers had normal or corrected-to-normal visual acuity and at least 20-arcsec stereoacuity as measured with the Randot® stereoacuity test (Precision Vision, La Salle, IL).

Apparatus

Scripts for stimulus presentation were executed on a G5 Power Macintosh using the Psychophysics Toolbox package (v. 3.0.8) (Brainard, 1997) for MATLAB (v. 7.4). Stimuli were presented on a pair of CRT monitors (ViewSonic G225f) arranged in a mirror stereoscope at a viewing distance of 0.6 m. The resolution of the monitors was 1280 × 960 pixels and the refresh rate was 100 Hz. At this resolution and viewing distance, each pixel subtended 1.77 arcmin. Observers used a chin rest to stabilize head position during testing and a mouse to make their responses.
Stimuli

Figure 2 shows the three types of stimuli used. All three stimuli were composed of a rectangular binocular occluder and a smaller monocular object. All stimuli were drawn on a gray background (46 cd/m²).

1) Bar stimulus. This stimulus was identical, including angular size, to that used by Nakayama and Shimojo (1990) in their experiment 2. It was composed of a solid, white binocular rectangle (138 × 418 arcmin) and a solid, white monocular bar (7.08 × 331 arcmin).

2) Line stimulus. This stimulus matched that used by Gillam et al. (2003) in terms of appearance. However, the dimensions of the occluder matched those used by Nakayama and Shimojo (1990). That is, the occluder was the same size as in the bar stimulus but was comprised of a thin white outline 1.77 arcmin wide and filled with white random-dots each subtending 1.77 × 1.77 arcmin (10% density). The monocular object in this case was a white line, 1.77 × 418 arcmin.

3) Disc stimulus. The occluder in this stimulus was the same as in the bar stimulus, but the monocular object was a white disc 7.08 arcmin in diameter (equal to the width of the bar in the bar stimulus and similar in size to the disc used in Gillam et al., 2003). We chose to use the original occluder rather than the modified occluder used by Gillam et al. (2003) to make a direct comparison with the condition assessed by Nakayama and Shimojo (1990).

For all stimuli, the occluding rectangle was positioned at zero disparity in the center of the display and a black fixation cross was placed at its center. The vertical lines of the cross were Nonius lines (3.5 × 23 arcmin) presented one to each eye, and the horizontal line (42.5 × 3.5 arcmin) was presented to both eyes with zero disparity. In each case the monocular test object (bar, line, or disc) was placed to the right of the center of the stimulus. It consisted of a solid, white binocular rectangle (138 × 418 arcmin) bisected by a small horizontal line subtending 35.4 × 3.5 arcmin and another, moveable, horizontal cursor line (35.4 × 3.5 arcmin). Observers moved the cursor along the ruler using a computer mouse (see Figure 2D). They were instructed to match the distance between the central bisection mark and the cursor to the perceived depth of the monocular object relative to the binocular rectangle. Observers moved the cursor above or below the central bisection mark to indicate that the monocular object was seen behind or in front of the rectangle, respectively. They also used this method to estimate depth in the binocular session. Observers were asked to fixate on the fixation cross while evaluating the depth in the stimulus and adjusting the ruler.

The three stimuli were shown in separate sessions whose order was randomized across observers. Each type of stimulus was shown 20 times for 240 trials per session (camouflage/occlusion × 6 occlusion widths × 20 repetitions). In each session, stimuli were presented in random order. Before each monocular stimulus session, observers completed a binocular session of 40 trials (4 disparities × 10 repetitions) in which the bar, line, or disc was binocular and had either crossed or uncrossed disparity. The purpose of these binocular sessions was twofold, to train the observers in using the estimation method and to generate comparison data for the monocular conditions.

Control experiments

We conducted three control experiments to (a) rule out the influence of eye movements and (b) assess whether the ability to estimate the lateral position of the
disc with respect to the rectangle differed compared to the line and bar (in light of the results of Experiment 1).

In the first control experiment (Experiment 1A), two observers (one naïve) performed a two-interval forced choice (2IFC) procedure to assess perceived depth with briefly presented stimuli (100 ms) (see Figure 3). Accurate binocular fixation was controlled and monitored with a fixation cross and Nonius lines. On each of the trials, one randomly chosen interval contained a monocular bar or disc stimulus (Figure 2A and C), with occlusion widths of 12.4, 15.9, or 19.5 arcmin, and the other interval contained a similar binocular stimulus presented with disparity that varied according to the method of constant stimuli (25 repetitions of seven disparities for each of the three occlusion widths). Observers were asked to indicate the interval in which the larger depth was perceived. A pilot, in which stimuli were defined by binocular disparity, was conducted to confirm the feasibility of the task.

In Experiment 1B we measured the just noticeable difference (JND) in the lateral position of a binocular bar or disc with respect to the rectangle. The binocular disc or bar was presented with zero disparity (so that it appeared at the same depth as the rectangle), and the distance between the rectangle and the outer edge of the bar or disc was varied. Five observers (three naïve) participated in a 2IFC discrimination procedure (same as in Experiment 1A). In each trial, in one randomly chosen interval the distance was set at 15.9 arcmin and in the other interval the distance was 15.9, 16.8, 17.7, 18.6, 19.5, 20.4, or 21.2 arcmin. Observers were asked to indicate the interval in which the larger distance was presented.

Experiment 1C also used the bar and the disc presented dichoptically with zero disparity but assessed the perceived magnitude of the lateral position judgments. The distance between the rectangle and the outer edge of the bar/disc was set at 12.4, 21.24, or 30.1 arcmin. Six observers (four naïve) indicated the perceived distance from the edge of the rectangle to the outer edge of the bar or disc using the virtual ruler procedure used in the main experiment.

Results and discussion

The left graph in Figure 4 shows the mean results for the sessions in which the bar, line, and disc were monocular. Depth estimates, expressed as equivalent disparity, are shown as a function of occlusion width, arrangement, and monocular object type. Distances estimated using the virtual ruler were converted to equivalent disparities using a standard equation relating on screen disparity and depth. The right side of the graph shows the data for the camouflage arrangement, while the left side of the graph shows the data for the camouflage arrangement. The black solid lines show the predicted disparity the estimates should follow according to occlusion geometry (or double-matching). Individual data for each observer are shown in Figure A1 in the Appendix.

For the bar and line stimuli in the occlusion arrangement, perceived depth increased with increasing occlusion width. On average, observers underestimated the depth relative to the predicted depth. Individual data (see Figure A1 in the Appendix) showed that two of the observers (LT and LW) followed the prediction...
closely, while the other four perceived less depth for both line and bar stimuli. These results were consistent with Nakayama and Shimojo (1990), who found that some observers’ depth estimates were smaller than the minimum predicted from the occlusion geometry. In the camouflage condition quantitative depth was on average seen for the line stimulus but not for the bar stimulus, in agreement with Gillam et al. (2003) and Nakayama and Shimojo (1990). On average, the bar was perceived slightly in front of the occluder and this percept was independent of occlusion width. Variability was noticeably higher in this condition. Some observers always perceived the bar to lie in front of the occluder, others always saw it behind or at the same depth; for some observers the percept changed from trial to trial (see Figure A1 in the Appendix). Observer DS perceived quantitative depth in the bar stimulus in the camouflage condition and observer LW did not see quantitative depth in the camouflage condition in any of the stimuli.

Following the logic of Gillam et al. (2003), the disc stimulus is the crucial test of the double-matching hypothesis. As shown in Figure 4A, the mean perceived depth of the disc increased as occlusion width increased. Five of the six observers perceived quantitative depth from the disc stimuli. Observer LW showed no increase in perceived depth with increasing occlusion width (see Figure A1 in the Appendix). In accordance with the existing literature, no quantitative depth was observed in the camouflage arrangement for the disc stimulus.

These observations were confirmed by statistical analysis. The data were analyzed separately for each observer and found to show the same trends as the mean data, so the mean data analysis is provided here. We first performed two-way repeated-measures ANOVAs with occlusion width and stimulus type as factors separately for the occlusion and the camouflage configurations. For the occlusion arrangement, the two-way ANOVA showed significant main effects of occlusion width, $F(5, 25) = 75.8, p < 0.001$, and stimulus type, $F(2, 10) = 9.2, p = 0.005$, as well as a significant interaction, $F(10, 50) = 4.8, p < 0.001$. The interaction is the result of the steeper increase in perceived depth for the bar and the line stimuli in comparison to the disc stimulus. In a post hoc analysis, we fitted regression lines to the data corresponding to each stimulus type. All slopes were found to be significantly different from 0 and 1 (bar, $b = 0.65 \pm 0.054$; line, $b = 0.67 \pm 0.048$; disc, $b = 0.35 \pm 0.056$). As expected the slopes of the bar and the line stimuli were almost identical, while the slope for the disc stimulus was significantly shallower.

In the camouflage arrangement the two-way ANOVA showed a significant effect of occlusion width, $F(5, 25) = 2.6, p = 0.049$, and significant effects of stimulus type, $F(2, 10) = 7.3, p = 0.01$, and the interaction of the two factors, $F(10, 50) = 6.46, p < 0.001$. The interaction is the result of the difference in the effect of occlusion width in the line stimulus and the bar and the disc stimuli. The slope for the line stimulus was found to be significantly different from zero and from the other two slopes (bar, $b = 0.65 \pm 0.09$; line, $b = 0.005 \pm 0.11$; disc, $b = -0.1 \pm 0.06$). The disc stimulus slope was also found to be different than zero, but it was very shallow and in the opposite direction from that predicted by geometry.

It could be argued that some of the observers did not actually see quantitative depth in the disc stimulus and based their depth judgments purely on the occlusion width. This is highly unlikely since in this case we would expect a similar strategy to be used in the camouflage condition, and that is clearly not the case. Additionally, the disparity estimates would be much smaller than observed since the estimated depth on the ruler would be approximately equal to the actual separation of the monocular feature from the rectangle and not to the depth predicted from that separation.

Another potential explanation for our results is that our disc stimulus was more prone to double-matching than the disc stimulus of Gillam et al. (2003), since our occluder and disc were both monochromatic. We evaluated this possibility by creating a disc stimulus with an occluder similar to that used by Gillam et al. (2003) (same occluder as in the line stimulus) and re-running the experiment with two observers (AC and IT). Both observers showed statistically significant quantitative depth in the occlusion arrangement with this new stimulus, confirming our original findings.

Finally, quantitative depth in our monocular configurations could have been based on alternating fixation. Observers could estimate depth from the difference in vergence angles created by alternating fixation between the occluder and the occluded object (possibly without awareness) or between the occluded object and the ruler (Enright, 1991). Since viewing time was unlimited, observers could potentially shift their gaze during a trial, despite instructions to maintain fixation. Further, the ruler was located 1.75° away from the edge of the occluder, which could potentially prompt the observers to break fixation. However, results of the first control experiment (Experiment 1A) suggest that eye movements were not an important factor. In this experiment stimuli were presented for only 100 ms so observers should not have had enough time to initiate vergence eye movements while the stimuli were visible (e.g., Yang, Bucci, & Kapoula, 2002). For each data set of this experiment the point of subjective equality was estimated as the 50% point in the psychometric function (cumulative normal) fit to the data. Given that the procedure compared binocular and monocular occlusion depths, these points represent
the “equivalent perceived disparity” of the monocular occlusion stimuli and are shown in Figure 5. It is clear that the perceived depth of the monocular features (bar and disc) increased significantly as a function of the occlusion width. Consequently, alternating fixation could be ruled out as an explanation of the results of Experiment 1. Interestingly, as in Experiment 1, perceived depth in Experiment 1A was less than the depth specified by the minimum constraint for both stimuli. This suggests that the underestimation observed in the monocular sessions of Experiment 1 (Figure 4A) was not due to the estimation method used.

Although in Experiment 1, the disc stimuli in the occlusion arrangement produced quantitative depth percepts, the magnitude of perceived depth was significantly smaller than that observed for the line and the bar stimuli. Interestingly, in the binocular session, when the bar, line, and disc were binocular and had either crossed or uncrossed disparity, there was a similar underestimation of depth in the disc stimuli as shown in the right graph of Figure 4 (individual data are shown in Figure A2 in the Appendix). The left side of the graph shows the crossed disparity condition and the right side shows the uncrossed disparity condition. In both cases, observers see less depth when viewing the disc stimulus than the other two stimuli. Repeated-measures two-way ANOVAs showed a significant effect for stimulus type, disparity and the interaction of the two factors for both uncrossed disparity—type, $F(2, 10) = 19.5, p < 0.001$; disparity, $F(1, 5) = 39.84, p < 0.005$; interaction, $F(2, 10) = 7.2, p = 0.01$—and crossed disparity—type, $F(2, 10) = 39.2, p < 0.001$; disparity, $F(1, 5) = 41.6, p < 0.005$; interaction, $F(2, 10) = 6.9, p = 0.012$. The significant interaction reflects the difference in the rate of change of depth with disparity between the disc stimulus and the other two stimuli. Regression analysis showed that slopes for the line and bar stimuli types for both uncrossed disparity—bar, $b = 1.05 \pm 0.19$; line, $b = 1.0 \pm 0.17$; disc, $b = 0.72 \pm 0.15$—and crossed disparity—bar, $b = 1.23 \pm 0.22$; line, $b = 1.21 \pm 0.21$; disc, $b = 0.84 \pm 0.22$—were significantly different from zero and not significantly different from one (the slope of the predicted depth). For the disc stimulus the slope was significantly different from one for uncrossed disparities but not for the crossed.

It is possible that quantitative depth in the monocular disc stimulus is smaller than in the other two conditions because this stimulus does not provide a reliable double-match. However, this does not explain why the binocular disc stimulus yields smaller perceived depth as well. A different account that encompasses both results appears more likely. Both depth from disparity and depth from monocular occlusion/camouflage in our stimuli require an accurate estimation of the relative position of the bar, line, or disc with respect to the rectangle (or the bar, line or disc in the other eye). Due to its small spatial extent it might be harder to estimate the position of the disc compared with the bar or the line. Combined with the visual system’s general bias towards smaller disparities this might result in smaller perceived depth for the disc. Two control experiments support this hypothesis. Experiment 1B showed that the mean JND for the position of the disc stimulus ($M = 1.84, SD = 0.28$ arcmin) was significantly larger, $t(4) = 7.5, p < 0.01$, than the mean JND for the position of the bar stimulus ($M = 0.85, SD = 0.17$ arcmin). Larger JND for the disc stimulus implies noisier estimates of its location relative to the rectangle. More noise in the disc condition could degrade depth estimates both from disparity and monocular occlusions compared to the bar stimulus.

In Experiment 1C, we directly compared the perceived lateral distance between the disc or bar and the edge of the rectangle using the virtual ruler (as in Experiment 1). Repeated-measures ANOVA confirmed that for all observers the perceived lateral distance of the disc or bar from the rectangle increased with physical distance, $F(2, 10) = 178.8, p < 0.001$, but was consistently larger for the bar than for the disc. $F(1, 5) = 29.38, p = 0.003$. The increase in estimated distance with true distance was greater for the bar than for the disc, interaction $F(2, 10) = 6.4, p = 0.016$. The finding that the perceived width of the “occlusion” region is smaller for the disc stimulus than for the bar stimulus maps well onto the smaller perceived depth in the monocular disc stimulus in Experiment 1 (Figure 4A).

Experiment 1C also showed that the degree of the underestimation of the position of the disc increased as its true lateral position increased (significant interaction in ANOVA). Thus, in the binocular case, the disparity of the disc could have been underestimated since in one...
eye the position of the disc would be underestimated more than in the other eye, yielding a smaller disparity between the two eyes than that of an identical line stimulus.

**Experiment 2**

This experiment was designed to probe the origin of depth in two-object arrangements using a different method from that used in Experiment 1. Tsirlin et al. (2011) reported that depth from monocular occlusions, but not from disparity, can be biased by a proximate binocular feature. They proposed that the presence of such depth biasing could serve as a litmus test for distinguishing depth originating from occlusion geometry versus depth originating from disparity.

There is some evidence in the literature that the depth of monocular objects in two-object arrangements can be biased by a nearby binocular feature. Nakayama and Shimojo (1990) noted that the disparity probe they used affected the perceived depth of the monocular object. Similarly, Hakkinen and Nyman (1996) found that an additional binocular surface placed above a monocular dot in a two-object arrangement biased its perceived location. However, neither of these studies asked if this bias was specific to monocular stimuli and so did not assess the influence on binocular versions of the stimuli. In this experiment we examined whether the depth of the monocular bar, line, or disc used in Experiment 1 could be biased by an adjacent binocular surface. We also assessed whether a similar bias was present when the bar, the line, or the disc was presented binocularly. If the bias existed only for the monocular versions of the stimuli, this would suggest that depth percepts in the two cases were based on different mechanisms, namely monocular occlusion geometry in the first case and disparity in the second.

**Methods**

The observers and apparatus were the same as in Experiment 1; however, only the occlusion arrangement was tested in this experiment. The occlusion stimuli (bar, line, and disc) were modified by adding a uniform gray surface behind the occluding rectangle in depth. The surface subtended $4.07^\circ \times 7.25^\circ$ and had a disparity of 17.7 arcmin. Both the occluder and the monocular features were located within the boundaries of the additional surface (see Figure 6). The observers were asked to ignore this additional surface and to perform the estimation task as they did in Experiment 1. As in Experiment 1, each session was preceded by a set of binocular trials in which the bar, the line, and the disc had uncrossed disparities of 7.08, 14.16, 21.24, or 28.32 arcmin. The experiment was completed in three sessions, one for each type of feature (bar, line, and disc), in random order.

**Results and discussion**

Results from this experiment, along with the results of Experiment 1, are presented in Figure 7. The mean depth estimates in Experiment 2 for monocular features (Figure 7A dashed lines) are substantially larger than those obtained in Experiment 1 (Figure 7A solid lines) for all feature types (this pattern was consistent across observers). However, there is little difference in the perceived depth of binocular features between the two experiments (Figure 7B). To quantify the effect of the additional surface on the perceived depth in Experiment 2, we computed the mean bias for stimuli with monocular and binocular features. The bias was computed for each observer individually first, by subtracting the depth estimates for each occlusion width or disparity and stimulus type obtained in Experiment 1 from the depth estimates obtained in Experiment 2, and then averaging these differences across occlusion width or disparity. The means for each stimulus type were then averaged across observers to obtain a final bias measurement for each condition. As shown in Figure 7C, the addition of a binocularly defined surface, behind the two-object arrangement, biases the perceived depth of all monocularly defined features but not the binocular features.
two-way repeated-measures ANOVA on the mean bias data with stimulus type and experiment type (monocular or binocular) as factors, showed a significant main effect of experiment, $F(1, 5) = 60, p < 0.001$, and no significant effects of stimulus type, $F(2, 10) = 0.3, p = 0.76$, or an interaction between the factors, $F(2, 10) = 0.3, p = 0.76$. In post hoc analyses, paired $t$-tests showed that for all three stimuli types the bias in the monocular and the binocular cases differed significantly, bar, $t(5) = 6.7, p < 0.005$; line, $t(5) = 3.4, p = 0.02$; disc, $t(5) = 4.6, p < 0.01$. In addition, one-sample $t$-tests showed that, for the binocular stimuli, the bias was not significantly different from zero for the bar and the line stimuli—bar, $t(5) = 0.6, p = 0.55$; line, $t(5) = 1.1, p = 0.3$—but was significantly different from zero for the disc—$t(5) = 3.7, p = 0.013$. All $t$-tests were confirmed with a nonparametric Wilcoxon test.

These data show that, as in Tsirlin et al. (2011), the depth of monocular but not binocular features can be easily biased in the direction of surrounding binocularly defined features. Importantly, all three types of stimuli, (bar, line, and disc) show a similar pattern of bias, suggesting that the same process, based on occlusion geometry, is used to assign depth.

**Computational modeling**

In this experiment, we used a computational approach to assess the origin of quantitative depth percepts in two-object arrangements. The double-matching hypothesis implies that a correspondence is made between the occluding and the occluded object. Any useful definition of a match implies a degree of similarity between matched images, which, in the simplest case, is a correlation. Moreover, it is widely believed that disparity detectors in V1 essentially perform a cross-correlation of the right and the left images (Banks, Gepshtein, & Landy, 2004; DeAngelis, Ohzawa, & Freeman, 1995). Although V1 disparity detectors were found to not be sufficient for stereopsis, they are believed to be the first step in computing disparity for the perception of depth as well as for maintaining fusion (Cumming & DeAngelis, 2001). To assess how well image correlation signals would support double-matching, in terms of reliability and consistency, we applied a windowed cross-correlation algorithm to the bar and the disc stimuli used in Experiment 1.

**Methods**

We used the sum of square differences (SSD) as a correlation metric (for review, see Scharstein & Szeliski, 2002). The disparity that generated the maximum correlation value (minimum SSD) was chosen as the estimated disparity for each pixel. In cases where more than one disparity generated the maximum correlation value (i.e., multiple identical peaks), three methods of disparity assignment were compared:

1. The mean of the peak correlation disparities was computed.
2. The median of the peak correlation disparities was computed.
3. The disparity closest to zero was selected (zero bias).

Based on this algorithm we obtained a disparity estimate for each pixel and formed disparity maps. A correlation strength metric for each pixel was computed by dividing the peak correlation value of this pixel by...
the largest correlation response in the whole image. A correlation reliability metric for each pixel was computed as the difference between the normalized peak correlation responses for this pixel and the next largest response (i.e., having two or more disparities with the maximum correlation value would generate a reliability of 0).

To accommodate disparity signals at different scales, we used three correlation window sizes, 4, 10 and 20 pixels, and two disparity ranges, ±20 pixels and ±30 pixels. We also applied the algorithm separately from the left to the right image and from the right to the left image (the direction of the computation has been shown to impact the detection of monocular areas; Egnal & Wildes, 2002). In total there were 36 different conditions for each stereogram (3 methods of depth estimation × 3 window sizes × 2 disparity ranges × 2 source images). We used four types of stereograms:

1. Monocular bar stimulus from Experiment 1
2. Monocular disc stimulus from Experiment 1
3. Binocular bar stimulus from Experiment 1
4. Monocular bar stimulus with a black rectangle and white bar on a gray background (opposite contrast polarity).

The monocular stimuli were all consistent with the geometry of occlusion and the right edge of the monocular bar and disc was 7, 12, or 18 pixels away from the occluder. The binocular bar stimulus was identical in all respects to the monocular bar stimulus except that the bar was presented to both eyes and had a disparity of 4, 8, or 12 pixels. The binocular stimulus was used to compare the performance of the algorithm on monocular occlusions with its performance on similar binocular areas. The opposite polarity bar stimulus was used to demonstrate the dependence of the algorithm’s response on other stimulus properties.

Results and discussion

Figure 8 shows the results of the simulations for right-to-left cross-correlation. We chose to show these results since the monocular features were in the right eye’s image and thus right-to-left application gave better depth estimates. The left-to-right case is discussed in following text. Estimated depth for monocular objects is highly dependent on the parameters of the cross-correlation algorithm (Figure 8 first and second rows). The estimates vary widely as a function of window size, disparity range, and disparity assignment method. No combination of parameters yields estimates close to those predicted by geometry (dashed black line). The combination that shows results most similar to our psychophysical data for both types of monocular objects (for same contrast stimuli)—un-
crossed depth that increases with occlusion width and is somewhat smaller than the predicted depth—combines the smallest correlation window (4 pixels) and the zero bias disparity assignment method (two top leftmost graphs, blue lines with square symbols). Other combinations give depth estimates that either do not increase with occlusion width or yield crossed or overly large disparities. Interestingly, in all cases, predicted depth in the monocular disc stimulus is smaller than that in the monocular bar stimulus.

On the other hand, depth from disparity in the binocular bar stimulus (Figure 8, third row) is very robust to algorithm parameters. All methods, window sizes and disparity ranges yield the same estimated disparities, which are virtually identical to the predicted values. Importantly, disparity signal strength and reliability were found to be low for the monocularly occluded features and high for the binocular bar under all conditions. Moreover, the left-to-right simulation provided a set of depth estimates for the monocular regions that differ from those shown in Figure 8. However, in the case of the binocular stimulus, the estimates for the left-to-right simulation are the same as those for the right-to-left. In fact many stereo algorithms use what is called a left-to-right check, comparing the output of the left-to-right to right-to-left applications, to identify monocularly occluded zones (Egnal & Wildes, 2002).

The ability of the correlation algorithm to match the monocular bar and disc to the occluder in the other eye is not surprising since cross-correlators are designed to respond to somewhat similar features in the two eyes (which is why they produce false matches in binocular areas). However, as our simulations show this signal is very noisy, unstable, and highly dependent on the parameters of the detectors. The algorithm is also very sensitive to stimulus properties. As shown in the bottom row of Figure 8, no veridical quantitative depth can be predicted by any parameter combinations we tried in the opposite contrast polarity bar stimulus. At the same time, observers perceive as much depth from this stimulus as they do from the same contrast polarity stimulus (Nakayama & Shimojo, 1990).

To conclude, the cross-correlation output for monocularly occluded regions in two-object arrangements is noisy and parameter dependent. This contrasts with the vivid and robust quantitative depth experienced by human observers for both the monocular bar and dot stimuli and suggests that estimating depth in monocular regions requires more sophisticated mechanisms than simple correlation-based disparity detectors. These mechanisms might include separate monocular occlusion detectors that identify monocularly occluded areas (Egnal & Wildes, 2002), as well as higher-level neurons that would integrate information from disparity detectors and monocular occlusion detectors to establish a depth map.
General discussion

The source of quantitative depth in two-object arrangements

The main focus of the present work is to identify the source of quantitative depth in two-object arrangements in order to understand the mechanisms involved in da Vinci stereopsis. Several explanations for quantitative depth in two-object arrangements, like that used by Nakayama and Shimojo (1990), have been proposed. We will consider each in turn and discuss the most likely scenario given the results of our current experiments as well as data from the literature.
Double-matching

Gillam et al. (2003) proposed that depth in two-object arrangements can be accounted for by conventional disparity, assuming the edge of the occluder is matched to both itself and an edge of the monocular object in the other eye. Evidence for double-matching is a matter of debate for stereopsis in general (see Howard & Rogers, 2002, pp. 127–137). More specifically for two-object arrangements a number of observations from our data, and the existing literature, argue against this hypothesis. First, Experiment 1 suggests that the failure to find quantitative depth in the monocular disc stimulus in Gillam et al. (2003) could be attributed to exceeding the range of da Vinci stereopsis. Following their logic, the fact that we obtained quantitative depth percepts using the disc stimulus argues against the double-matching hypothesis (see also Zannoli & Mamassian, 2011).

Second, it is well known that binocular disparities up to 1°–2° can yield quantitative depth percepts (for review, see Wilcox & Allison, 2009). Consequently, it is reasonable to assume that if depth perception in two-object arrangements relies on disparity it should be possible to obtain quantitative depth percepts for a wide range of occlusion widths. Thus, the very limited range of occlusion widths over which quantitative depth in two-object arrangement can be perceived (up to only 15–30 arcmin depending on the stimulus) is inconsistent with disparity-based depth.

Third, our Experiment 1 and Nakayama and Shimojo (1990) also showed no quantitative depth in the camouflage condition for the bar and the disc stimuli. If double-matching takes place in two-object arrangements, quantitative depth should be seen in the camouflage condition since crossed and uncrossed disparities are treated (relatively) equally by the visual system. Quantitative depth was seen by five out of six observers in the camouflage condition with the line stimulus, which could be interpreted as a sign of double-matching. However, given the other arguments in this section, the results of Experiment 2 as well as the fact that occlusion geometry also predicts quantitative depth from camouflage, it does not seem likely that double-matching is the source of depth percepts in the line stimulus.

Fourth, in Experiment 2 we demonstrated that the perceived depth of all three monocular objects in two-object arrangements could be biased towards a far binocular surface, while the perceived depth of identical binocular objects was resilient to biasing. It could be argued that due to double-matching the disparity signal in two-object arrangements is weak and hence is more prone to influences from nearby binocular surfaces. However, Tsirlin et al. (2011) showed that even in stimuli with weak disparity signals most observers do not exhibit biasing.

Finally, we showed that model disparity detectors produce a weak, noisy, and parameter-dependent depth signal for monocular regions in two-object arrangements. This pattern is different from the stable, robust response to similar binocular stimuli. This discrepancy suggests that, although quantitative depth can sometimes arise in monocular regions due to fortuitous parameter choices in simple disparity detector models, the vivid depth percepts obtained in such stimuli are more likely to depend on a more sophisticated set of mechanisms.

In sum, the weight of the evidence from our experiments and the literature does not favor a double-matching account of depth percepts from two-object arrangements.

Sequential matching

Even in the absence of double-matching, depth in two-object arrangements could still be based on sequential stereopsis (Enright, 1991). When fixated on the occluder, the edges of the occluder would be fused and the monocular object would be unmatched. When fixated on the monocular object, it could be matched with the nearest edge of the occluder, leaving the occluder itself unmatched. The observer could estimate the relative disparity between the two objects based on change in the vergence angle. Control Experiment 1A was designed to address this issue, and it showed that quantitative depth is still seen in two-object arrangements when fixation is controlled and exposure durations are too brief to complete a vergence eye movement (Figure 5). The amount of depth perceived by both observers in Control Experiment 1A was similar (and sometimes slightly exceeded) that estimated in Experiment 1. Hence we can rule out this explanation.

Occlusion geometry

The explanation for da Vinci stereopsis offered by Nakayama and Shimojo (1990) was based on occlusion geometry. They argued that the visual system could use the constraints on the minimum depth imposed by occlusion geometry (see Figure 1B) to assign depth to the monocular objects in such stimuli. Data from Experiment 1 and Experiment 2 are consistent with this explanation as are the results of several other studies that used two-object arrangements (Hakkinen & Nymann, 1996; Nakayama & Shimojo, 1990). Experiment 1 shows that, for the majority of observers, quantitative depth could be perceived in the disc stimulus, in which double-matching is unlikely. Experiment 2 provides new evidence that depth percepts in two-object arrangements are more likely to stem from occlusion constraints than from disparity since the depth of the
monocular features in these arrangements can be strongly biased by nearby disparity unlike the depth of similar binocular features.

Moreover, there is a growing body of evidence that the visual system uses the constraints imposed by occlusion geometry to assign quantitative depth to illusory occluding surfaces (Anderson, 1994; Gillam & Nakayama, 1999; Grove & Gillam, 2007; Pianta & Gillam, 2003; Sachter & Gillam, 2007; Tsirlin et al., 2010). Quantitative depth based purely on occlusion geometry has also been found in moving stimuli (Brooks & Gillam, 2006; Shimojo et al., 1988). Thus if the constraints are used by the visual system in two types of monocular occlusion stimuli, parsimony argues that the same mechanism is used with other types of occlusion stimuli such as two-object arrangements.

Two findings seem to be at odds with the monocular geometry explanation. First, depth estimates in Experiment 1 were smaller than those predicted by the minimum depth constraint for four out of six of our observers. The only explanation that we can offer for this discrepancy is that the mechanisms estimating depth from monocular occlusion geometry are not as precise as disparity detectors. Note that this underestimation is also an issue for a disparity-based explanation. Another issue is the absence of quantitative depth in camouflage arrangements. However, this is a more serious problem for simultaneous or sequential double-matching accounts, as discussed in more detail in the following text.

Why does camouflage not work?

Experiment 1 confirmed that quantitative depth is not perceived in camouflage arrangements when the monocular object is a bar or a disc. Observers perceived these monocular objects slightly in front, slightly behind, or at the same depth with the occluder, but depth percepts did not vary consistently with occlusion width. Quantitative depth from camouflage configurations was perceived in the line condition in Experiment 1 (and in Gillam et al., 2003); however, the variability was greater in this case than in the equivalent occlusion condition (see Figure 4). If depth in da Vinci stereopsis is indeed based on double-matching (or another disparity-based process), the absence of quantitative depth in the camouflage condition with bar and disc stimuli is hard to explain. Gillam et al. (2003) proposed that in the original bar stimuli used by Nakayama and Shimojo (1990), the bar was shorter than the occluding rectangle and thus had strong relative size cue indicating it was positioned behind the occluder, which could have hindered depth perception in the camouflage condition.

This explanation is unlikely since if we make the bar (or the disc) binocular, it is perceived in front of the rectangle and, more importantly, exhibits precise quantitative depth as was shown in Experiment 1, even though the same size cues are in place. Moreover, as shown in Figure 9B when the monocular bar is extended to the full length of the occluder, the percept is very similar to the original stimulus shown in Figure 9A. We know that crossed and uncrossed disparities are treated (relatively) equally by the visual system, hence the absence of quantitative depth in the camouflage condition poses a serious problem for the double-matching hypothesis.

If depth in two-object arrangements is based on occlusion geometry, camouflage configurations should also yield quantitative depth, unless these phenomena are not represented equally by the visual system (unlike crossed and uncrossed disparities). Nakayama and Shimojo (1990) proposed that this asymmetry makes sense ecologically since camouflage in two-object arrangements would occur only rarely in natural viewing since it requires an almost perfect match between the color and luminance of the occluder and the occluded object. Thus it is possible that the visual system simply cannot represent monocular camouflage. Occlusion, on the other hand, frequently occurs at depth discontinuities, and it is likely that cortical mechanisms have developed to resolve depth in these areas.

Following this reasoning, it is possible that camouflage arrangements are reinterpreted as occlusion arrangements by creation of an illusory occluding edge on the other side (opposite to the binocular rectangle) of the monocular object (see also Assee & Qian, 2007). Our observers reported seeing such edges in some camouflage stimuli, and there are numerous examples in the literature of monocular occlusions promoting the perception of illusory occluding contours and surfaces (e.g., Anderson, 1994; Gillam & Nakayama, 1999; Tsirlin et al., 2010). In this case, the occlusion width should have no effect on the perceived depth of the monocular object since its depth is estimated relative to the illusory edge (which presumably does not change its location relative to the monocular object). This explanation, however, does not account for the presence of quantitative depth in the line stimulus in Experiment 1 and in Gillam et al. (2003) experiments. It is possible that depth percepts in monocular occlusion stimuli depend on the size of the occluded object and the occluder. The disc and the bar had the same width in our experiments while the line was much thinner (as in Gillam et al., 2003); this hypothesis could account for the differences in percepts found in these experiments but requires further investigation.

Interestingly, quantitative depth in camouflage arrangements is readily perceived when the monocular
object is extended towards the occluder such that no gap is left between the two objects and an “extrusion” is formed as shown in Figure 9C. Gillam et al. (2003) and Cook and Gillam (2004) suggested that cyclopean T- or L-junctions might be one of the factors affecting depth perception in these extrusion stimuli in comparison with conventional two-object arrangements. Cyclopean T- or L-junctions, they proposed, might serve as additional indicators of occlusion or camouflage relationships since T- and L-junctions in general play a role in figure-ground assignment in monocular and binocular images (Anderson, 1997; Anderson & Julesz, 1995). If this is the case, however, a similar, clear percept of crossed depth should be obtained from the stimuli shown in Figure 9D and Figure 9E, where cyclopean T- and L-junctions are also present. However, in these cases the monocular regions are perceived by most observers to lie beyond the occluder.

It is more likely that the depth seen in extrusion stimuli is due to conventional disparity. That is, the right edge of the rectangle in the right image is matched to the right edge of the figure in the left image to produce the resulting depth profile. Importantly, this is an instance of one-to-one matching and not double-matching. To illustrate this point, a simplified, equivalent contour stereogram is shown in Figure 9F, which produces the same depth profile as Figure 9C. Consistent with the above explanation, depth is not readily seen in Figure 9D since double-matching would be required here to obtain depth since the right edge of the binocular rectangle is visible. In Figure 9E one-to-one matching is complicated since the corresponding edges in the right and left eyes have opposite contrast polarity.

Cook and Gillam (2004) reported finding accurate quantitative depth in figure-eight intrusion stereograms, which are conceptually similar to the extrusion stimuli shown here. They suggested that this percept in their stimulus relies on occlusion geometry since the shape of the intruding surface was seen as flat rather than the concave shape predicted from one-to-one matching. It is possible, however, that depth magnitude in their stimulus was established based on matching, while the perceived flat shape of the surface could have resulted from the interpretation of occlusion relationships and the contours of the intrusion. This idea was supported by computational simulations (similar to those presented in this article) with the Cook and Gillam (2004) stimuli, that showed that accurate and reliable depth in extrusion or intrusion stimuli can be computed by simple disparity detectors (Tsirlin, Allison, & Wilcox, 2012).

To conclude, it is not yet clear why some instances of the camouflage configuration do not produce quantitative depth perception. The ecological explanation proposed by Nakayama and Shimojo (1990) together with the reinterpretation of the camouflage arrangement as an instance of occlusion by an illusory surface seem to be the most viable alternatives.

**Mechanisms underlying da Vinci stereopsis**

The evidence presented here suggests that quantitative depth in two-object arrangements relies on
monocular occlusion geometry rather than disparity. Importantly, taken together with previous findings, this result suggests that there is a single mechanism that underlies quantitative and qualitative depth perception in da Vinci stereopsis. Our psychophysical and computational experiments also suggest that da Vinci stereopsis cannot simply be a by-product of the activity of low-level disparity detectors. Instead it is likely to require a network of sophisticated detectors involving binocular cells specifically tuned to identify monocularly occluded areas. The importance of explicit detection of monocularly occluded areas has been stressed in the computer vision literature (for review, see Egnal & Wildes, 2002) and biologically inspired models of stereopsis have recently started to address this issue (Assee & Qian, 2007; Hayashi, Maeda, Shimojo, & Tachi, 2004; Watanabe & Fukushima, 1999). Further, as our experiments show, depth estimates computed for monocularly occluded areas could be based on occlusion geometry and governed by the minimum depth constraint. However, none of the existing models explicitly use the width of the monocular zones in order to assign depth to monocular areas or illusory occluders. The viability of this strategy requires further investigation.

Acknowledgments

We would like to thank Barbara Gillam, Kevin MacKenzie, Adam Reeves, and an anonymous reviewer for insightful comments on our experiments. This research was supported by NSERC grants to L. M. Wilcox and R. S. Allison and a CGS NSERC scholarship to I. Tsirlin.

Commercial relationships: none.
Corresponding author: Inna Tsirlin.
Email: itsirlin@yorku.ca.
Address: Centre for Vision Research, Department of Psychology, York University, Toronto, Ontario, Canada.

Footnotes

1 It appears that, unlike Experiment 1, in Experiment 1A the perceived depth of the disk stimulus was not much smaller than that of the line stimulus. However, in Experiment 1 depth was estimated using the same virtual ruler for both stimuli, hence depth estimates were directly comparable. In Experiment 1A the comparison was between a monocular and a binocular version of the stimulus (bar or disk).

Therefore, the properties causing the difficulty estimating the location of the disc in comparison to the bar, as reported in Experiments 1B and 1C, were present in both the target and the probe. This explains why the results of Experiment 1A would not show the depth magnitude differences between the bar and the disc. However, we can still compare the slopes of the functions relating perceived depth magnitude versus occlusion width. As in Experiment 1 the slope of the function for the disc stimulus is shallower than that for the bar stimulus (Figure 5 vs. Figure 4A).

2 When computing the bias for the binocular features, we used only the data corresponding to the disparity values that were used in Experiment 1, 7.08 and 14.16 arcmin, as shown in Figure 7B.

References


Appendix

Figure A1. Individual data from Experiment 1 with monocular stimuli. Estimated disparity is shown as a function of stimulus type and occlusion width. Data for the occlusion arrangement are shown on the right and the data for the camouflage are shown on the left of each graph. The black solid lines show the theoretical disparity. Error bars show ±1 SEM.

Figure A2. Individual data from Experiment 1 with binocular stimuli. Estimated disparity is shown as a function of stimulus type and stimulus disparity. Data for the uncrossed disparity are shown on the right and the data for crossed disparity are shown on the left of each graph. The black solid lines show the theoretical disparity. Error bars show ±1 SEM.