



Sensory loss due to object formation

Pascal Mamassian, Marina Zannoli

► To cite this version:

Pascal Mamassian, Marina Zannoli. Sensory loss due to object formation. Vision Research, 2020, 10.1016/j.visres.2020.05.005 . hal-02876784

HAL Id: hal-02876784

<https://cnrs.hal.science/hal-02876784>

Submitted on 21 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Sensory loss due to object formation

Pascal Mamassian^{a,*}, Marina Zannoli^b

^a Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

^b Facebook Reality Labs, Redmond, WA 98052, USA

ARTICLE INFO

Keywords:

Visual perception
3D
Binocular disparity
Grouping
Uncertainty
Motion
Binding

ABSTRACT

The precision to locate individual features in depth can often be improved by integrating information over space. However, this integration can sometimes be extremely detrimental, as for example in the case of the Westheimer-McKee phenomenon where features are grouped to form an object. We replicate here the known loss of precision in this phenomenon and document an additional loss of accuracy. These detrimental effects are still present when the object is elicited by other principles of organization, including a cross-modal auditory cue. Similar effects of object formation are found on lateral motion sensitivity. We then present a simple probabilistic model based on the integration of estimated depth within an object and propagation of object mean depth and uncertainty back to the elementary features of the object. This propagation of object uncertainty is a hitherto underestimated side-effect of object formation.

1. Introduction

When visual information is processed in the visual system, more and more sophisticated representations are generated, from elementary features processed in the retina (Nassi & Callaway, 2009) up to complex objects represented in the anterior temporal cortex (DiCarlo, Zoccolan, & Rust, 2012; Peelen & Caramazza, 2012). However, these computations come at a cost, namely that of discarding local information along the way to produce more abstract representations. Very early on in the investigation of visual phenomena, principles of perceptual organization have been proposed to generate these more abstract representations (Wagemans et al., 2012a). Grouping is commonly believed to be a beneficial feature of perception because it helps discard the variability within an object to extract the gist or skeleton of the object (Buhmann, Malik, & Perona, 1999; Feldman & Singh, 2006; Alvarez, 2011). However, not all variability is noise, and in some cases, grouping can be detrimental. For instance, Liu, Jacobs, and Basri (1999) have shown that when the central part of an object is hidden behind an occluder, it is more difficult to discriminate the depth of the two parts on either side of the occluder when there is strong evidence that the two parts belong to the same object (see also Hou, Lu, Zhou, & Liu, 2006). We present here more evidence that object formation can sometimes be detrimental for perception, and offer a simple model to account for these effects.

The integration of visual information within an object relies on a proper segmentation of the object from other objects. The segmentation of objects is especially important during the processing of depth

information, such as that coming from binocular and motion cues, when we estimate the three-dimensional structure of objects (for a review of depth cues, see Welchman, 2016). For motion, it is well accepted that some form of segmentation occurs before motion information is integrated within segmented objects. Clearly, it would make no sense to integrate motion information across two distinct objects because these different objects are free to move in different directions and different speeds (Braddick, 1993; Masson & Perrinet, 2012). In comparison to the motion literature, the importance of segmentation in scene analysis is encountered less often in studies on binocular vision. A notable exception is the work of Mitchison (1988) who clearly advocated that binocular disparities needed to be segmented into different groups before being interpreted. This author went on to argue that the visual system was attempting to estimate planar surfaces for each segmented group, as if one was trying to approximate the scene as a set of piecewise planar surfaces. Surface segmentation is also present in the comprehensive model of Grossberg (1987) and in the recent work of Cammack and Harris (2016) that is more extensively reviewed in the discussion.

Apart from these few exceptions, the segmentation issue is blatantly absent in some major reviews of stereopsis (Poggio & Poggio, 1984; Cumming & DeAngelis, 2001; Parker, 2007; Blake & Wilson, 2011). The reason is maybe an over-emphasis on the correspondence problem where one is interested in explaining which elementary features in the left eye's image are matched to which ones in the right eye's image. The correspondence problem became a central issue in psychophysics and

* Corresponding author at: Laboratoire des Systèmes Perceptifs (CNRS UMR 8248), Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France.

E-mail address: pascal.mamassian@ens.fr (P. Mamassian).

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Frontal View							
Plan View							
Depth Sensitivity	good	bad	bad	good	bad	good	good

Fig. 1. Westheimer-McKee phenomenon. The sensitivity to discriminate the relative depth of two vertical lines depends on the spatial configuration. In this set of examples, the top row illustrates what the observer sees from a cyclopean viewpoint, the second row is a view from above, and the third row is a simplified summary of how well the two vertical lines can be separated in depth. This depth sensitivity varies greatly even though the vertical lines have always the same location and size. See text for details.

computational modelling when random dot stereograms (RDS) became popular (Julesz, 1971), because these stimuli demonstrated that stereopsis was possible even when there was no monocular form present. In other words, binocular depth could be perceived even when the scene was not pre-segmented into individual objects. This observation has maybe led to the belief that segmentation is not important for stereopsis. In contrast however, binocular disparities are appreciated as a critical cue for object segmentation (Harris & Wilcox, 2009).

The interest for the correspondence problem has completely set aside binocular phenomena where the correspondence is obvious. One of the most striking phenomena in this respect is what we may call the *Westheimer-McKee phenomenon*. Westheimer (1979) noted that when two vertical lines placed at different depths were connected so as to form a square, observers were no longer able to tell which line was in front of the other (Fig. 1a and b). McKee (1983) confirmed this observation and also found a decrease in sensitivity when the two vertical lines were just connected by a single horizontal segment so as to form an “H” figure (Fig. 1c). This latter finding rules out the hypothesis that closure is a necessary constraint to get the loss in depth sensitivity.

It is important to note that the Westheimer-McKee phenomenon is very sensitive to minor changes in the spatial configuration of the stimulus. McKee (1983) noticed that a small break in the connections of the vertical lines, so as to form an open square bracket figure on the left of a vertical line, was sufficient to bring back sensitivity almost to the level where there were only two isolated lines (Fig. 1d). In contrast, Mitchison and Westheimer (1984) found that sensitivity was almost as bad as the one obtained with the slanted square when the stimulus was cut in the middle, so as to display open and closed square brackets facing each other (Fig. 1e). In this case, the end-points of the brackets were assigned disparities in such a way that both the open and the closed square brackets were contained in a single slanted plane. But when the brackets were placed in separate fronto-parallel planes, sensitivity was good again (Fig. 1f). Finally, Westheimer (1979) noted that when the square brackets point outwards, keeping them in separate fronto-parallel planes, sensitivity was also very good (Fig. 1g). Even though McKee (1983) found an earlier description of the loss of stereoacuity due to closure in a long monograph by Werner (1937), it is fair to give to Gerald Westheimer and Suzanne McKee the credit for the first critical psychophysical demonstrations of this phenomenon.

The striking Westheimer-McKee phenomenon can be appreciated when one displays multiple copies of the stimulus side by side (Mamassian, 2008). Fig. 2A illustrates two object configurations, both composed of vertical lines where pairs of adjacent lines are connected together so as to form rectangles. The vertical lines are assigned one of two possible depths, with two consecutive lines having the same depth, the next two the other depth, and so on. When the two lines of a rectangle have different depths, one obtains small planes slanted in depth that look like vertical blinds alternating left and right in slant. In contrast, when the lines of a rectangle have the same depth, one obtains small fronto-parallel planes alternating front and back. Fig. 2B shows these two configurations in a stereo pair, one above the white cross, one

below it. Note that the vertical lines for the two object configurations are perfectly aligned in each eye’s image, therefore the binocular disparity information to discriminate which vertical lines are in front is the same in both configurations. And yet, as we measure more precisely later in this paper, the fronto-parallel configuration leads to a better depth percept than the slanted rectangles.

By measuring the estimated depths for supra-threshold stimuli in two configurations (Fig. 3), Deas and Wilcox (2014) generalized the Westheimer-McKee phenomenon that was originally described near depth threshold. The authors asked their participants to estimate the perceived depth between the central two vertical lines of their stimuli by adjusting the distance between their thumb and index fingers on a touch-sensitive sensor. They found that the estimated depth of vertical lines that belong to a closed object was reduced relative to the lines presented in isolation. In contrast, the estimated depth of the central vertical lines when they belonged to two distinct objects was similar to the lines presented in isolation. We note however that one of the two objects was again an object slanted in depth just like in the previous configuration, so the perceived depth in this configuration is between the edges of two different objects, one fronto-parallel and one slanted (Fig. 3). This mixture makes a quantitative interpretation of the results more complex. Deas & Wilcox also manipulated the grouping strength of the two vertical lines and observed systematic changes in estimated depth. These results led them to conclude that closure was a critical constraint in stereopsis, in contrast to the previous result of McKee (1983) reviewed above (see again Fig. 1c).

The Westheimer-McKee phenomenon is interesting in that it is an example where grouping is detrimental. While several models have been proposed to mimic grouping principles (Wagemans et al., 2012b), these models emphasize the benefits of merging multiple features into coherent objects. In this report, we present a model that can account for both the beneficial and the detrimental effects of perceptual grouping. We argue that the detrimental effect on the perception of the features of an object results from a propagation of the uncertainty in the representation of the object.

In the following, we report a series of psychophysical experiments to measure the effects of object formation on both the precision and the accuracy of depth perception of vertical lines. Detrimental effects are revealed when object formation is obtained by physically connecting the vertical lines, grouping the lines by similarity of color, or using some cross-modal congruency between depths and sounds. We also generalize some of these effects to motion perception. We then propose a simple model that can account for the deterioration in performance due to object formation, and rule out alternative interpretations. Even though the main experiments are based on a stereoscopic judgment of elementary binocular features, the results and model described here can have general implications for sensory object formation.

2. Experiment 1: Effect of object formation on stereoacuity

The purpose of the first experiment was to study the Westheimer-

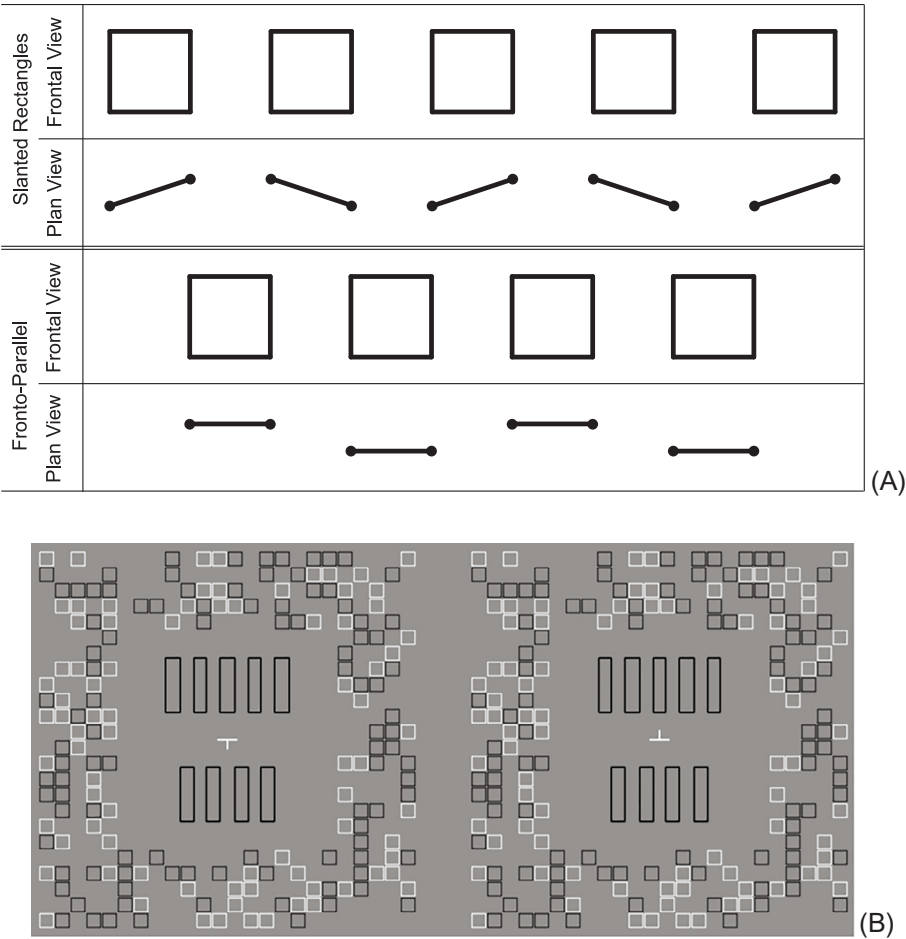


Fig. 2. Duplication of the Westheimer-McKee phenomenon. (A) Illustration of the spatial arrangement in two configurations. The vertical lines in both configurations are at the same locations, therefore from a cyclopean viewpoint, the observer sees a similar set of rectangles side by side in both configurations. The difference between the two configurations is whether vertical lines within each rectangle have the same or different depths. (B) Stereogram. The left and right images are stereo pairs that can be free-fused. Above the nonius lines are vertical lines connected by small horizontal segments so as to form planes slanted in depth. Below the nonius lines are vertical lines connected by segments so as to form fronto-parallel planes alternating in front and behind fixation. Note that the vertical lines above and below the nonius are perfectly aligned, so the binocular disparity information is identical in both configurations. In spite of resting on the same binocular information, the slant of the slanted planes above is more difficult to perceive than the depth between the fronto-parallel planes below. Figure replicated from Mamassian (2008).

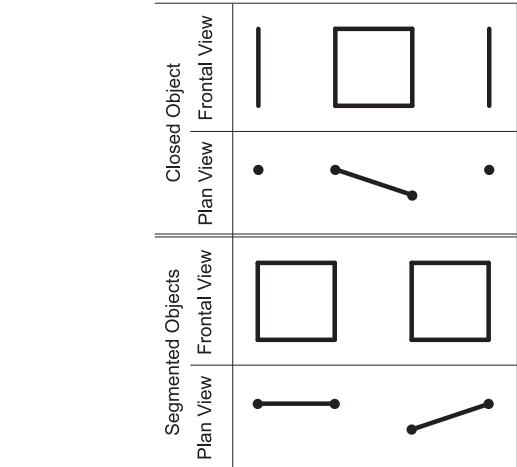


Fig. 3. Stimulus configuration used by Deas and Wilcox (2014). In the closed object configuration, the two central vertical lines are grouped into a slanted object. In the segmented objects configuration, one object is fronto-parallel and the other is slanted. Observers were asked to estimate the depth extent between the two central vertical lines.

McKee phenomenon with multiple objects (Fig. 2). The task of participants was always to judge the depth order between two vertical lines that were either presented on either side of a fixation point. We were interested in both the precision and the accuracy with which the lines were perceived in depth when the lines were displayed in isolation or belonged to an object (Fig. 4). In the first part of the experiment that targeted depth precision, we contrasted a configuration where the

vertical lines belonged to two distinct fronto-parallel objects and a configuration where the lines belonged to a slanted object. In the second part of the experiment that targeted depth accuracy, the vertical lines belonged to two separate objects, both slanted in the same direction. In contrast to the work of Deas and Wilcox (2014), fronto-parallel and slanted objects were thus measured separately.

2.1. Methods

2.1.1. Participants

Five participants took part in Experiment 1 after giving their informed consent. All participants had normal or corrected to normal visual acuity and normal stereoacuity. Stereoacuity was assessed by performance in the first condition of the experiment (two vertical lines at different depths on either side of fixation) with an inclusion criterion set to stereoacuity thresholds of 5 arcmin or less. The studies were conducted while the authors were both at University Paris Descartes and were part of a project that was approved by the CERES ethics committee (Conseil d'évaluation éthique pour les recherches en santé) of University Paris Descartes.

2.1.2. Apparatus

A modified Wheatstone stereoscope was used in the stereopsis and cross-modal experiments. Stimuli were displayed on a split-screen 21" SONY Trinitron CRT GDM-F520 monitor set at a resolution of 1280 × 1024 pixels and running at 75 Hz. The monitor was gamma-corrected with a Minolta chromameter CS-100A. Stimuli were generated and presented with the PsychToolbox (Brainard, 1997; Pelli, 1997). Head movements were restrained by a forehead chinrest, and the monitor was viewed at an optical distance of 89 cm (the sum of the

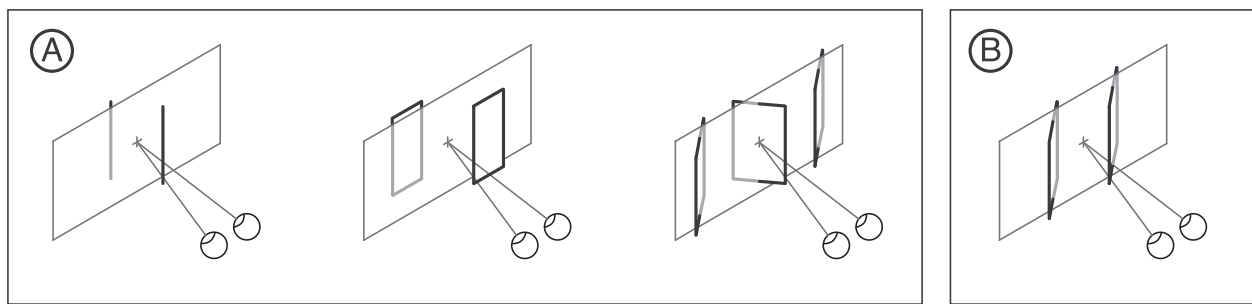


Fig. 4. Stimulus configurations to study the effects of object formation on stereo sensitivity and biases. In all the experiments, observers had to pay attention to the two vertical lines closer to fixation and decide whether it was the one on the right or the one on the left that was in front. (A) In Experiment 1A, three stimulus configurations were contrasted. The stimuli consisted of two, four, or six vertical lines placed either behind or in front of fixation. When four or six vertical lines were presented, small horizontal segments were added at their extremities so as to form either fronto-parallel or slanted planes. (B) In Experiment 1B, four vertical lines were presented grouped in pairs so as to form two parallel slanted planes.

optical distances across the mirrors).

2.1.3. Stimuli

Horizontal binocular disparities of a line corresponded to a lateral displacement in one eye's image relative to the other eye's image. Antialiasing was used to allow subpixel positioning. To help maintain vergence, a frame composed of black and white small squares was continuously presented (see Fig. 2B) and nonius lines were shown before each trial. The stimuli consisted of black vertical lines on a mid-grey background. The lines were 1 degree of visual angle long, 2 arcmin thick, and separated horizontally by 15 arcmin. All disparities reported here are relative disparities between the two vertical lines, that is twice the absolute disparity of each line relative to fixation.

In Experiment 1A, there were three stimulus configurations (Fig. 4A). The two central lines were presented with different disparities within a range that was different depending on the configuration. In the “baseline” configuration, only two vertical lines were presented on either side of fixation. In the “fronto-parallel” configuration, two extra lines were presented, one on each side of the lines shown in the baseline configuration. These extra lines were presented at the same disparity as their neighboring line and connected with it with small horizontal segments so as to form two fronto-parallel rectangles at different depths. In the “slanted-rectangles” configuration, two extra lines were presented on each side of the lines shown in the fronto-parallel configuration. These extra lines were presented at a disparity opposite to that of their neighboring line. The resulting six line were connected in adjacent pairs by small horizontal segments so as to form three vertical slanted rectangles of alternating signs (left–right–left or right–left–right).

In Experiment 1B, there were always four vertical lines grouped together so as to form two slanted rectangles (Fig. 4B). The two central lines were presented with different disparities within a range that was different depending on the displayed slants. The slant was identical in both rectangles (same amplitude and same sign) and was chosen across trials out of five values. The slants were obtained by adding a disparity to the outer edge of each rectangle (i.e. the edge on the other side than the side containing the edge to discriminate in depth). This added disparity could be -1 , -0.5 , 0 , 0.5 , or 1 arcmin, thereby generating two configurations where slants were oriented to the left, one fronto-parallel, and two to the right.

2.1.4. Procedure

The task of the observers was to attend to the two lines that were on either side of the fixation point and to report which one (left/right) was perceived closer to them. The method of constant stimuli was used to build psychometric functions relating probability of perceiving the right line in front to binocular disparity. The range of horizontal disparities presented to the observers was predefined for each condition and based

on pilot data. In Experiment 1A, the psychometric functions were based on 32 repeated trials for each of 8 disparity levels in each of the three configurations. The resulting 768 trials per observer were run intermixed in 8 blocks of trials. In Experiment 1B, the psychometric functions were based on 24 repeated trials for each of 8 disparity levels in each of the five object slants. The resulting 960 trials per observer were run intermixed in 8 blocks of trials. All participants first ran Experiment 1A and then Experiment 1B.

Nonius lines were presented before each stimulus to encourage good vergence. The nonius lines consisted of a white ‘+’ sign with the horizontal part visible in both eyes and the vertical part split between the two eyes (an example is shown in Fig. 2B). Observers looked at the nonius lines for at least 500 msec and pressed a key when satisfied that the two vertical parts of the ‘+’ sign were aligned. The stimulus was then presented for 800 msec and replaced by the nonius lines. Observers then reported which line they perceived in front by pressing either the left or the right arrow key on the keyboard. No feedback was ever provided in any of the experiments reported in this paper.

2.1.5. Statistical analysis

The effect of experimental conditions on the probability of perceiving one of the two lines in front was analysed with a generalised linear mixed effects model in MATLAB R2017b (*fitglme.m*). In Experiment 1A, the model included binocular disparity and experimental condition (“baseline”, “fronto-parallel”, and “slanted-rectangles”) as fixed effects and participant as random effect. Condition was considered a categorical factor. A probit link function was indicated to connect horizontal disparity between the two lines to the probability that the right line was perceived in front. The analysis can reveal a main effect of disparity that is interpreted as a non-zero slope of the psychometric function. The analysis can also reveal the effect of the experimental condition on the slope and point of subjective equality (PSE) of the psychometric function. This type of analysis considers between-observer variation as a random factor and has the advantage of higher statistical power compared to traditional analysis using a two-level approach where one first estimates the parameters of the psychometric function for each observer and then tests for significant differences of these parameters across observers (Moscatelli, Mezzetti, & Lacquaniti, 2012). Overall sample set effects were tested with an ANOVA (*anova.m*) and post hoc tests via contrast matrices (*coefTest.m*). Mixed-effects hierarchical models are inherently conservative and circumvent the need for multiple comparisons corrections (Gelman, Hill, & Yajima, 2012).

2.2. Results

In Experiment 1A, participants viewed binocularly two vertical lines presented side by side and were prompted to report which one

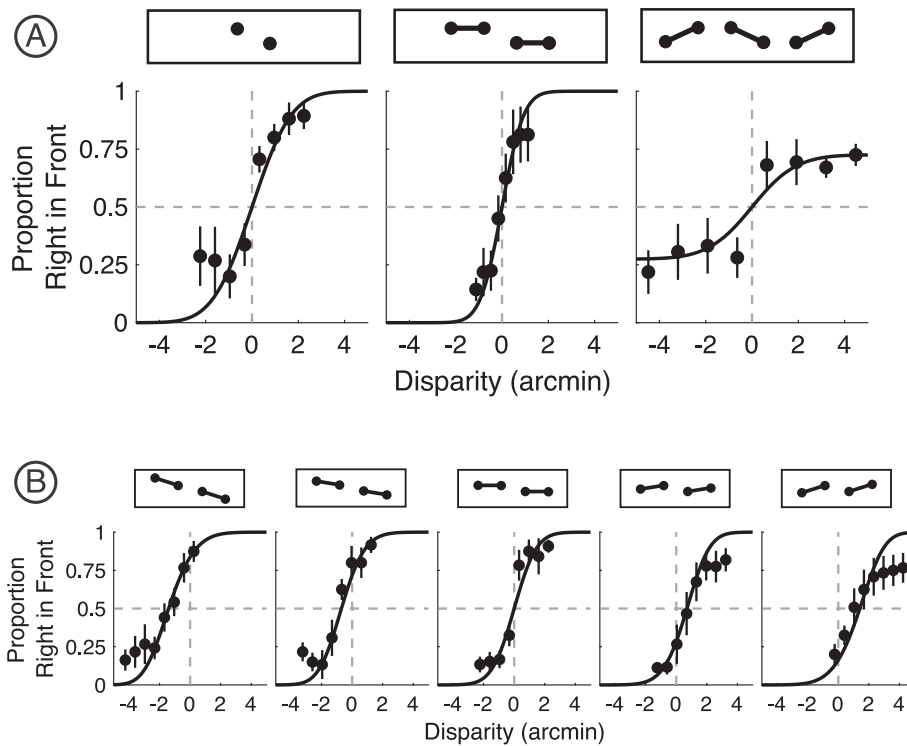


Fig. 5. Influence of object formation on stereoacuity in Experiment 1. (A) Results for the five observers in Experiment 1A. When only two vertical lines are presented (illustrated by two dots in the panel showing a plan view), deciding which one is in front is very easy (steep psychometric function). When these two lines are the central edges of two fronto-parallel surfaces, performance is slightly improved (steeper psychometric function). When the two lines are connected so as to form a slanted rectangle, depth discrimination is dramatically impaired, even though planes slanted in opposite direction on either side could potentially be used as a contrasting cue. Solid lines are the best fit of the propagation of uncertainty model (see Fig. 13 and accompanying text for details). (B) Results for the same observers in Experiment 1B. The lines to be discriminated in depth now belong to two parallel slanted planes. As slant magnitude increases, there is an increasing bias as visible by a shift of the point of subjective equality. There is also a small decrease in sensitivity as shown by a decrease in the slope of the psychometric function. The solid lines are predictions from the model whose parameters were already estimated in Experiment 1A (the model is therefore not fitted to these data). Symbols are data averaged across five observers, error bars are standard errors across observers.

appeared closer to them (Fig. 4A). This setup is a version of the standard Howard-Dolman test to measure stereoacuity (Howard & Rogers, 1995). Performance in this task can be summarized by fitting a psychometric function (a cumulative Gaussian) to the probability of seeing the right line in front as a function of disparity. The disparity threshold to reach 75 percent correct averaged across observers was 1.22 arcmin (Fig. 5A). Equivalently, depth discriminability can be estimated for a given disparity between the left and right lines. For a disparity of 2 arcmin (corresponding to an absolute disparity of 1 arcmin for each line), performance reached 83.6 percent correct on average (95% confidence interval (95CI) obtained from 1000 bootstrap samples: [80.9, 86.5]).

In the fronto-parallel configuration, performance increased so that for a disparity of 2 arcmin, the average performance rose to 97.9 percent correct (95CI: [96.4, 98.8]). As expected from the Westheimer-McKee phenomenon, performance decreased dramatically in the slanted-rectangles configuration. In this configuration, performance quickly reached a plateau beyond which increasing disparity was not accompanied by an increase in performance (Fig. 5A). Sensitivity in this third configuration led to performance for a disparity of 2 arcmin that averaged 69.8 percent correct (95CI: [64.5, 72.9]), a performance significantly worse than that obtained when the lines were presented in isolation or grouped to form fronto-parallel objects.

These observations were confirmed by our statistical analysis. We ran a generalized linear mixed effects model with disparity and condition (“baseline”, “fronto-parallel”, and “slanted-rectangles”) as fixed effects and participant as random effect. We found an overall main effect of disparity $F(1, 3448) = 6.87$, $p = 0.009$, and a main effect of condition on the slope of the psychometric function $F(2, 3448) = 4.67$, $p = 0.009$. Post-hoc tests revealed that the slope of the psychometric function in the “slanted-rectangles” condition was significantly worse than the one in the “baseline” condition ($F(1, 3448) = 9.18$, $p = 0.002$). Even though there was a trend for the slope of the psychometric function in the “fronto-parallel” condition to be better than the one in the “baseline” condition, this difference did not reach significance ($F(1, 3448) = 3.58$, $p = 0.058$). The slope of the psychometric function in the “slanted-rectangles” condition was also

significantly worse than the one in the “fronto-parallel” condition ($F(1, 3448) = 6.65$, $p = 0.010$).

In Experiment 1B, participants viewed two slanted planes presented side by side and were asked to judge which of the two edges near the central fixation point was closer to them (Fig. 4B). Increasing the slant of the planes generated a bias in the depth discrimination. The bias was such that a slant to the right (with the right edge further back than the left edge) shifted the left edge of the object on the right backwards (and the right edge of the object on the left forwards). Increasing the slant of the planes also resulted in a worse depth discriminability, as indicated by a shallower psychometric function.

These observations were backed up by statistical analyses. We ran a generalized linear mixed effects model with disparity and condition (absolute value of the slant of the stimulus, 3 values) as fixed effects and participant as random effect. We found an overall main effect of disparity $F(1, 4302) = 8.01$, $p = 0.005$, a main effect of condition on the PSE of the psychometric function $F(1, 4302) = 14.6$, $p < 0.001$, and a main effect of condition on the slope of the psychometric function $F(1, 4302) = 5.10$, $p = 0.024$.

2.3. Discussion

The results of Experiment 1A are consistent with previous reports indicating that grouping can affect the perceived depth of a single object (Westheimer, 1979; McKee, 1983; Mitchison & Westheimer, 1984; Deas & Wilcox, 2014). In some of these earlier studies, different stimuli were compared that did not contain the same binocular information. For instance, in the stimulus composed of open and closed square brackets (see again Fig. 1e and f), the end points of the brackets contain themselves some disparities that could help or hinder the overall interpretation. In our experiment, we endeavoured to offer the same amount of disparity information across all three conditions. The vertical lines were placed horizontally at regular intervals and the grouping information came from the small horizontal segments that do not contain additional horizontal disparities. If anything, the slanted-rectangles configuration that led to worse performance contained more information because the side planes whose slant was opposite to that of

the central plane could generate a contrasting contextual effect (van Ee, Banks, & Backus, 1999).

In Experiment 1B, observers were presented with two parallel slanted planes and had to discriminate the depth of the two central edges (Fig. 4B). Increasing the slant of the planes introduced large and consistent biases in perceived depth. Just like for Experiment 1A, these results are somewhat surprising because the vertical lines to be discriminated in depth have the same binocular disparity information across all slant conditions, and yet changing the object slant is sufficient to create large biases.

It is important to note that these large effects on binocular sensitivity and biases cannot be accounted for by any traditional model of stereopsis (Ohzawa, DeAngelis, & Freeman, 1997; Tsai & Victor, 2003; Banks, Gepshtein, & Landy, 2004; Cao & Grossberg, 2005; Hibbard, 2007; Haefner & Cumming, 2008; Goutcher, Connolly, & Hibbard, 2018). The failure of contemporary models to account for the reported effects comes from the fact that these models attempt to solve the correspondence problem with no reference to the object that the local features belong to (an exception might be the model of Cao & Grossberg, 2005). In a later section, we propose a simple model that can account for the effects described here.

3. Experiment 2: Effect of object size on stereoacuity

The purpose of the second experiment was two-fold. First, we were interested in manipulating the strength of the grouping cue and for this purpose we increased the horizontal separation between the vertical lines in our stimuli. We reasoned that increasing the distance between lines should decrease the perception of a unified object. Second, we looked at the genericity of our results by testing a different grouping cue, namely the similarity of color of neighboring lines.

In Experiment 2A, we determine the extent to which grouping depends on the size of the object. In this experiment, we restrict the analysis to fronto-parallel and slanted-rectangles configurations. In the “fronto-parallel” configuration, stimuli consisted of four fronto-parallel rectangles alternating in depth, whereas in the “slanted-rectangles” configuration, stimuli consisted of five slanted rectangles alternating in 3D orientation. As previously, participants had to judge the relative depth on the most central vertical lines. We manipulated the width of the object by choosing three horizontal spacings of the vertical lines.

In Experiment 2B, we determine whether object closure is critical to obtain the loss of binocular sensitivity. We chose a different grouping cue, namely the color similarity of the features. Stimuli now consisted of vertical lines that were either white or black on a grey background (Fig. 6). Two consecutive lines of the same color seemed to be grouped because of the Gestalt principle of similarity (Wagemans et al., 2012a).

3.1. Methods

3.1.1. Participants

Three participants took part in Experiment 2A and two participants in Experiment 2B.

3.1.2. Apparatus

The same modified Wheatstone stereoscope that was used for Experiment 1 was used here.

3.1.3. Stimuli

The stimuli consisted of black vertical lines on a mid-grey background. The lines were 1 degree of visual angle long, 2 arcmin thick, and regularly spaced horizontally. The horizontal separation was called “object size” and was chosen among three levels, 10, 15 or 20 arcmin. There were two experimental conditions, a “fronto-parallel” and a “slanted-rectangles” configurations similar to those used in Experiment 1A. In an attempt to boost the effects of grouping, we increased the number of objects in the scene relative to Experiment 1. The “fronto-

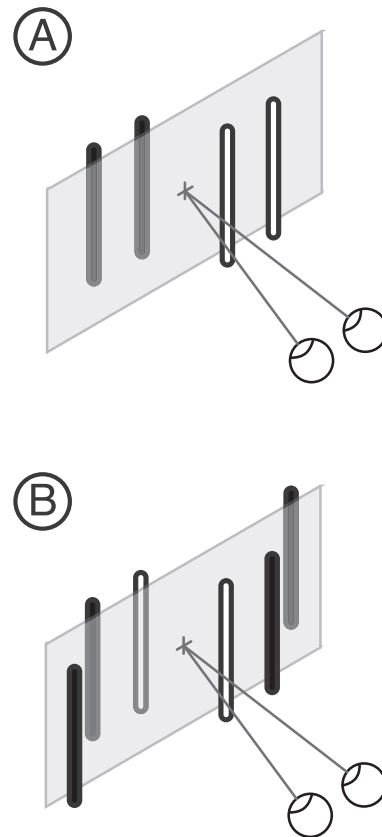


Fig. 6. Illustration of stimulus configurations used in Experiment 2B. Object formation was induced by color similarity: pairs of white (or black) lines tended to be grouped together. In the actual experiment, there were more lines than illustrated here. (A) The top illustration shows an example of fronto-parallel configuration. (B) The bottom illustration shows a slanted-rectangles configuration.

parallel” configuration consisted of four fronto-parallel rectangles alternating in depth, whereas the “slanted-rectangles” configuration consisted of five slanted rectangles alternating in 3D orientation.

In Experiment 2A, grouping was achieved with small horizontal segments just like in Experiment 1 (see examples in Fig. 2B). In Experiment 2B, grouping was achieved by color similarity, the vertical lines being either black or white (2 blacks followed by 2 whites and so on; see illustration in Fig. 6).

3.1.4. Procedure

After screening participants for normal stereoacuity using the standard Howard-Dolman test from experiment 1A, each observer was tested in two experimental conditions (“fronto-parallel” and “slanted-rectangles” only, no “baseline”) and three object sizes. Just like in Experiment 1, the task of the observer was to attend to the two lines that were on either side of the fixation point and to report which one (left/right) was perceived closer to them. Psychometric functions were obtained with Accelerated Stochastic Approximation (ASA) staircases (Kesten, 1958). Two staircases of 30 trials each were interleaved, repeated twice (so 120 trials total), for each of the two experimental conditions and for each of the three object sizes for each observer. Each experimental condition and object size were run in separate blocks of trials.

3.2. Results

We ran a generalized linear mixed effects model with disparity, condition (“fronto-parallel” and “slanted-rectangles”) and size (3 levels)

as fixed effects and participant as random effect. Condition was considered to be a categorical variable whereas size was considered a continuous variable.

In Experiment 2A, grouping was achieved by connecting pairs of vertical lines with small horizontal segments. We found a main effect of disparity $F(1, 1657) = 5.91$, $p = 0.015$, and a main effect of condition on the slope of the psychometric function $F(1, 1657) = 7.55$, $p = 0.006$. However, there was no significant interaction between the size and the condition on the slope of the psychometric function $F(1, 1657) = 0.15$, $p = 0.70$.

In Experiment 2B, grouping was achieved by displaying pairs of vertical lines with the same contrast (either black or white lines on a grey background). We found a main effect of disparity $F(1, 1432) = 82.6$, $p < 0.001$, and a main effect of condition on the slope of the psychometric function $F(1, 1432) = 27.2$, $p < 0.001$. Again, there was no significant interaction between the size and the condition on the slope of the psychometric function $F(1, 1432) = 0.22$, $p = 0.64$.

3.3. Discussion

We replicated the main effects of Experiment 1A with stimuli of different sizes and grouping that was provided by color similarity rather than closure. In both Experiments 2A and 2B, there were strong losses of depth sensitivity in the slanted-rectangles configuration as compared to the fronto-parallel condition. However, we failed to find an effect of object size on the strength of this loss in sensitivity, at least within the range over which we manipulated the horizontal separation between the vertical lines. It is as if once an object is segmented, it has the same effect on the sensitivity of its features, irrespective of the object shape and size.

Different grouping cues have different strengths (Wagemans et al., 2012a). Later in this paper, we describe a model to account for the effect of grouping on stereoacuity. This model includes a grouping saliency index that could be used to compare the strength of these different cues, for instance grouping by color similarity rather than closure. In the next section, we ask whether grouping can also be achieved through another sense.

4. Experiment 3: Cross-modal coupling

Instead of grouping lines with a visual cue, we attempted to group them thanks to a cue presented in a different sensory modality. To achieve this cross-modal coupling, we presented the vertical lines in a sequence, one line at a time instead of all of them simultaneously (Fig. 7). Visual stimuli consisted of three lines, the first two presented with identical disparity, and the last line with a different disparity (placing that line either front or back relative to the first two). Participants had to judge the relative depth of the last two lines: was the third line in front of or behind the second line? Each vertical line was paired with a pure tone that could have either a low or high pitch. We manipulated the compatibility of the three sound pitches with the three disparities. In the compatible configuration, each depth value was coupled with its own pitch. For instance, far disparity was associated with low pitch and near disparity with high pitch. In the incompatible configuration, the first line was coupled with one pitch and the last two with another pitch. This resulted in the last two lines that had different disparities to have the same sound associated to them. We reasoned that this sound association would contribute to group the last two lines into a common object, and thus predicted that depth discrimination would be worse in this incompatible condition compared to the compatible configuration.

4.1. Methods

4.1.1. Participants

Five participants with normal visual and hearing acuities took part

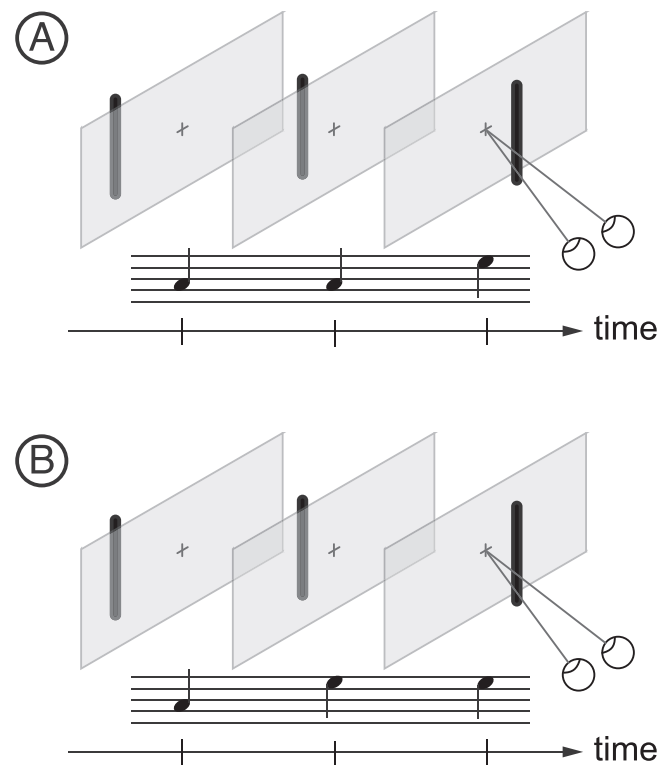


Fig. 7. Stimulus configurations in Experiment 3. Object formation was elicited cross-modally where binocular lines were presented one by one, each one simultaneously with a sound of identical or different pitch. The first two lines are presented at the same depth (back or front) and the third line at a different depth (front or back). (A) In the compatible association, the first two sounds have the same pitch and the third sound a different pitch, similarly to the sequence of depths. (B) In the incompatible association, the first sound has a different pitch than the last two sounds. Therefore, in this configuration, there is a change of depth that is not accompanied by a change of sound. In addition to these two configurations, there was also a baseline condition where the same pitch was played for each of the three visual stimuli.

in this experiment.

4.1.2. Apparatus

The same modified Wheatstone stereoscope that was used for Experiment 1 was used here.

4.1.3. Stimuli

Visual stimuli consisted of black vertical lines on a mid-grey background. The lines were 1 degree of visual angle long, 2 arcmin thick, and separated horizontally by 8 arcmin. The three lines were presented one after the other, from left to right or vice versa randomly across trials. A small jitter (max. of ± 0.8 arcmin) was added to the horizontal position of each line independently. The three lines were distributed around a disparity pedestal (max. of ± 2 arcmin) so that the absolute disparity of the last line was uninformative for the task. Each line was presented for 150 ms with an interstimulus interval of 50 ms. Sounds were pure tones of either 440 or 660 Hz that were presented in synchrony with each line.

4.1.4. Procedure

The first two lines were always presented at one particular disparity, and the last line at another disparity. The task of the observer was to report whether the last line was in front or behind the before-last line. There were three stimulus configurations depending on what sounds were played with the three lines. In the “baseline” condition, all three sounds had the same pitch, either low or high. In the “compatible”

condition, the first two sounds had the same pitch and the last sound a different pitch. In the “incompatible” condition, the first sound had one pitch and the last two sounds a different pitch. Therefore, in this “incompatible” condition, disparities were inconsistently paired with specific pitch, and the last two lines could appear grouped together because the same sound was played. The three stimulus configurations were interleaved within each of the 3 blocks of trials. The method of constant stimuli was used to build psychometric functions, with 8 horizontal disparities and 24 repeats per level (so 192 trials per psychometric function).

4.2. Results

We ran a generalized linear mixed effects model with disparity, condition (“baseline”, “compatible” and “incompatible”) as fixed effects and participant as random effect. We found a main effect of disparity $F(1, 2863) = 228, p < 0.001$, and a main effect of condition on the slope of the psychometric function $F(2, 2863) = 4.68, p = 0.009$. Post-hoc tests revealed that the slope of the psychometric function in the “incompatible” condition was significantly worse than the one in the “baseline” condition ($F(1, 2863) = 5.68, p = 0.017$), and the slope of the psychometric function in the “incompatible” condition was also significantly worse than the one in the “compatible” condition ($F(1, 2863) = 6.74, p = 0.009$; see Fig. 8), but there were no significant differences between the slopes of the psychometric functions in the “compatible” and “baseline” conditions ($F(1, 2863) = 0.057, p = 0.811$).

4.3. Discussion

In this experiment, we replicated the main effect of object grouping on depth sensitivity using grouping across sensory modalities. This time the object was defined cross-modally following the principle that an object produced a sound that had a constant pitch, and a different object had a different pitch. Therefore, playing the same pitch for lines that had different depths created the illusion that these lines belonged to a common object and made depth discrimination worse. Importantly, in this experiment just like the others in this paper, binocular information was identical across conditions. Whether the lines belonged to a single or two cross-modal objects changed the sensitivity to discriminate them in depth.

In other experiments not reported here, we varied the number of events in a trial. Instead of using a sequence of three lines, we tested conditions with only two lines (the minimum to make a depth judgment between them) and six lines (Zannoli, 2012). We replicated the results found here, namely that depth discrimination was worse when the pitch

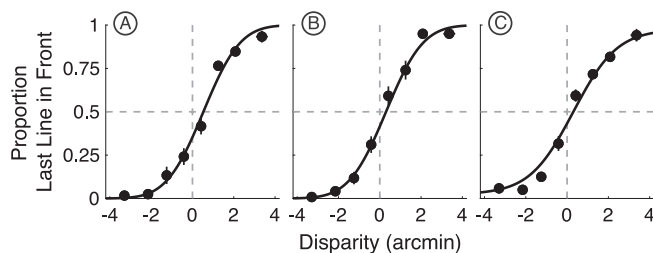


Fig. 8. Influence of object formation from auditory grouping on stereo acuity. A sequence of three lines was presented such that the first two were at the same absolute disparity and the last line either in front or behind of the first two. A sound with a high or low pitch was played simultaneously with each line. (A) In the baseline condition, the three sounds had the same pitch. (B) In the compatible condition, each line disparity was associated with a specific sound pitch. (C) In the incompatible condition, the first sound was played at one pitch and the last two at the other pitch. The inconsistency between disparity and pitch made depth discrimination worse. Error bars are standard errors across observers.

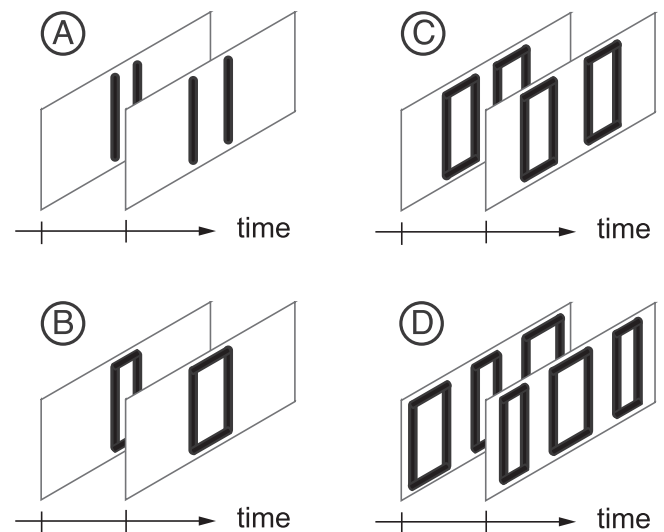


Fig. 9. Stimulus configurations to study the effects of object formation on motion sensitivity in Experiment 4. The effect of object formation was studied in a two-frame motion where the observer had to pay attention to the two vertical lines closer to the center and decide whether they moved towards (contraction) or away (expansion) from each other. Only expansion motion is illustrated here, contraction is obtained by inverting the order of the frames. The four stimulus configurations consisted in “2-lines” (A), “1-rectangle” (B), “2-rectangles” (C), and “3-rectangles” (D).

was the same across depths rather than different. In addition, this difficulty increased further when six lines were presented rather than two, suggesting that grouping was more salient when there was a consistent mapping between depth and pitch across multiple events. To further investigate the effect of grouping on stereoacuity, it would be interesting to manipulate the strength of the grouping on a trial-by-trial basis. This could be done by varying the proportion of compatible and incompatible pairings within a sequence.

5. Experiment 4: Effect of object formation on motion sensitivity

The purpose of the fourth experiment was to test whether object formation had similar effects in the motion domain. We replicated Experiments 1A and 2A by replacing binocular disparity with lateral motion in a two-frame stimulus. A set of vertical lines was placed with a regular horizontal spacing. In this experiment, we consider four different configurations by varying the number of lines and how these lines are connected (Fig. 9). Participants had to pay attention to the two central lines and discriminate contraction (lines get closer to each other) from expansion motion. In a baseline configuration, only two lines were presented. In the “1-rectangle” configuration, the two lines were connected with small segments so as to form a rectangle that became “thinner” (contraction) or “fatter” (expansion). In the “2-rectangles” configuration, two rigid objects moved toward or away from each other. Finally, the “3-rectangles” configuration was similar to the “1-rectangle” with two other deforming rectangles on either side.

5.1. Methods

5.1.1. Participants

Nine participants with normal visual acuity took part in this experiment.

5.1.2. Apparatus

Stimuli were displayed on the same monitor that was used for Experiment 1 running at 75 Hz and seen from a 57 cm viewing distance.

5.1.3. Stimuli

The stimuli consisted of black vertical lines on a mid-grey background and were viewed monocularly. They were 1 degree of visual angle long, 3 arcmin thick, and separated horizontally by 30 arcmin. Antialiasing was used to allow subpixel positioning. Motion was induced by presenting two consecutive frames of durations 400 and 40 msec respectively, without any inter-frame delay. There were four stimulus configurations. In the “2-lines” and the “1-rectangle” configurations, there were only two lines that moved in opposite directions, leftwards or rightwards (Fig. 9A and 9B). In the “2-rectangles” configuration, there were four lines, the first two lines moved in one direction and the next two in the opposite direction (Fig. 9C). Finally, in the “3-rectangles” configuration, there were six lines. The first line moved in one direction, the next two in the opposite direction, the next two back in the first direction, and the last one in the opposite direction (Fig. 9D).

To remove the possibility that participants use the absolute location of some stimulus feature instead of perceived motion, a small random horizontal jitter of the whole scene was introduced (uniform distribution in ± 2.0 arcmin range) across trials. Likewise, to avoid that participants use the aspect ratio of the rectangles as a proxy for perceived motion, a small random vertical jitter of the line heights was introduced (uniform distribution in $\pm 10\%$ range) across trials.

5.1.4. Procedure

The task of the observer was to attend to the two lines that were on either side of the fixation point and to report whether they were moving inward or outward. Psychometric functions were obtained with ASA staircases (Kesten, 1958). Two staircases of 20 trials each were interleaved, repeated six times (so 240 trials total), for each of the four experimental conditions for each observer. Experimental conditions were interleaved within each of the six blocks of trials.

5.2. Results

We ran a generalized linear mixed effects model with displacement, condition (“2-lines”, “1-rectangle”, “2-rectangles”, and “3-rectangles”) as fixed effects and participant as random effect. Line displacements were chosen from a staircase procedure. Even though line displacements are binned for illustrative purposes in Fig. 10, the raw values were entered in the statistical analysis. We found a main effect of displacement $F(1, 8625) = 108$, $p < 0.001$, and a main effect of condition on the slope of the psychometric function $F(3, 8625) = 4.20$, $p = 0.006$. Post-hoc tests revealed that the slope of the psychometric function in the “1-rectangle” condition was significantly worse than the

one in the “2-lines” condition ($F(1, 8625) = 5.42$, $p = 0.020$), and the slope of the psychometric function in the “3-rectangle” condition was also significantly worse than the one in the “2-lines” condition ($F(1, 8625) = 12.4$, $p < 0.001$), but there were no significant differences between the slopes of the psychometric functions in the “2-rectangle” and “2-lines” conditions ($F(1, 8625) = 1.82$, $p = 0.177$). Even though there was a trend for the slope of the psychometric function in the “3-rectangle” condition to be worse than the one in the “2-rectangle” condition, this difference did not reach significance ($F(1, 8625) = 3.25$, $p = 0.072$).

5.3. Discussion

Two conditions, the “2-lines” and the “2-rectangles” configurations, led to good sensitivity to discriminate contraction from expansion motion. In these conditions, pairs of lines moving in the same direction were grouped together with small horizontal segments, thereby creating the impression of two rigid rectangles moving inwards or outwards. In contrast, the other two conditions, namely the “1-rectangle” and the “3-rectangles” configurations, led to worse performance. In these conditions, the stimulus displayed either one or three non-rigid rectangles, that were either contracting or expanding. Note that these deformations would also be consistent with rigid planes rotating in depth about their vertical axis. Whether it is the non-rigidity of the object or its more complex rotation in depth that is at the origin of the worse performance cannot be answered here, but our results clearly show that motion within an object is challenging.

Other authors have shown how object formation can improve motion discriminability when the motions are similar (Verghese & Stone, 1996; Verghese & McKee, 2006). Our results go beyond these previous findings by showing that object formation can also hinder performance when the motions are dissimilar. In short, motion between objects is more visible than motion within an object.

6. Experiment 5: Effect of local object formation on global objects

So far, all of our experiments asked participants to attend to local features that belonged to different object configurations. In our final experiment, we looked at the perceived global slant of a group of objects. These objects could be slanted rectangles (Fig. 11A) or fronto-parallel rectangles (Fig. 11B). Importantly, the centers of the rectangles in both configurations were placed at the same locations within the global slanted plane (Fig. 12B). This property entails that if each rectangle is summarized by its averaged depth, both configurations should look the same. Therefore, this last experiment can be seen as a test as to whether observers are just estimating the mean depth of segmented objects.

6.1. Methods

6.1.1. Participants

In this experiment, there were only two participants with normal visual acuity and normal stereo acuity.

6.1.2. Apparatus

The same modified Wheatstone stereoscope that was used for Experiment 1 was used here.

6.1.3. Stimuli

The stimuli consisted of 8 vertical lines aligned horizontally with a regular spacing. There were three possible separations between the vertical lines, either 10, 15, or 20 arcmin. Each pair of adjacent lines was grouped with small horizontal segments so as to form small vertical rectangles similar to those used in Experiment 1.

Binocular disparities were added in the same direction and same amount every other line. In the “staircase” configuration, the first two

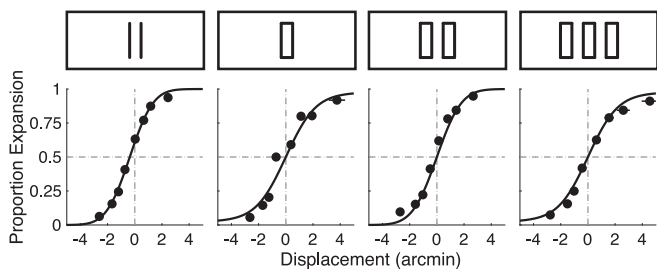


Fig. 10. Influence of object formation on motion acuity. Sensitivity was good for the “2-lines” configuration (first panel) and “2-rectangles” configuration (third panel). Performance was worse when the two lines whose motion was to be discriminated belonged to a single object, as this is the case for the “1-rectangle” configuration (second panel) and “3-rectangles” configuration (fourth panel). The stimuli presented to 9 observers were chosen from staircases and are binned here in 8 bins for illustration purposes. Horizontal and vertical error bars (often smaller than symbol size) are standard errors across observers. A cumulative Gaussian was fitted to the 2-line configuration and the smooth curve in the other three configurations show the fits of the propagation of uncertainty model (see Fig. 13 and accompanying text for details).

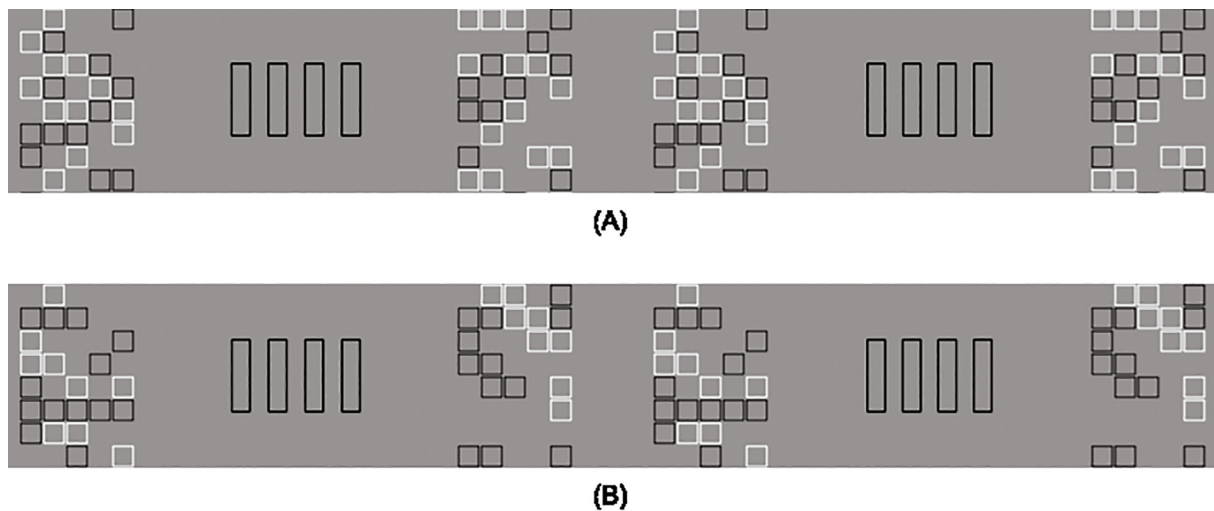


Fig. 11. Perception of global slant. Eight vertical lines were presented so that every other line displayed a constant depth increment. Stereo pairs are shown to illustrate two stimulus configurations that differed in the way these lines were connected. (A) In the “roof” configuration, the consecutive pairs of lines that were connected had different depths. (B) In the “staircase” configuration, the consecutive pairs of lines that were connected had the same depth. See Fig. 12B for an illustration of what these stimuli look like from above. Observers had to decide whether the two rectangles on the left were closer or further away than the two rectangles on the right, a global slant judgment.

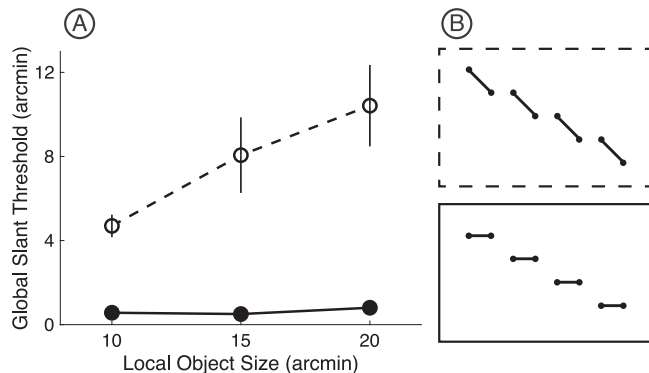


Fig. 12. Perception of global slant. (A) Results are shown in filled symbols for the “staircase” configuration and open symbols for the “roof”. Error bars are standard errors across participants. (B) Illustrations of the configurations are shown viewed from above. By construction, the two configurations look similar from a cyclopean viewpoint, and the mean depth for each of the four rectangles is the same in both configurations.

lines were given a disparity value of $-\frac{3}{2}\delta$, the next two lines had disparities incremented by δ (i.e. $-\frac{1}{2}\delta$), the next two lines again incremented by δ (i.e. $+\frac{1}{2}\delta$), and the last two lines again incremented by δ (i.e. $+\frac{3}{2}\delta$) (Fig. 12B). In the “roof” configuration, the first line had a disparity -2δ , the next two lines had disparities incremented by δ (i.e. $-\delta$), the next two lines again incremented by δ (i.e. 0), the next two lines again incremented by δ (i.e. $+\delta$), and the last line again incremented by δ (i.e. $+2\delta$). Because the procedure to add disparities between pairs of lines was identical in both configurations, this ensured that the global slant of the plane passing through the centers of the four rectangles was identical.

6.1.4. Procedure

Observers had to decide whether the two objects on the left were closer or further away than the two objects on the right, a global slant judgment. Psychometric functions were obtained with ASA staircases (Kesten, 1958). Two staircases of 30 trials each were interleaved, repeated 4 times (so 120 trials total), for each of the 3 object sizes and each of the 2 experimental conditions for each observer. Experimental conditions were interleaved within each block of trials, whereas each

object size was run in separate blocks.

6.2. Results

The global slant of the “staircase” stimulus was consistently perceived better than that of the “roof” stimulus (Fig. 12A). We ran a generalized linear mixed effects model with disparity, condition (“staircase” and “roof”) and size (3 levels) as fixed effects and participant as random effect. We found a main effect of disparity $F(1, 1432) = 6.45$, $p = 0.011$, and a main effect of condition on the slope of the psychometric function $F(1, 1432) = 8.20$, $p = 0.004$. There was a trend for an effect of size on the slope of the psychometric function that did not reach significance $F(1, 1432) = 3.18$, $p = 0.075$.

6.3. Discussion

Two configurations were created using vertical lines such that every other line presented a disparity increment. Consecutive pairs of lines were connected together to display vertically oriented rectangles, the difference between the two configurations being whether the lines had the same or different disparities (Fig. 12B). When the lines of a grouped pair had the same disparity, the stimulus was perceived as fronto-parallel rectangles that looked like a “staircase” if viewed from above. When the lines of a pair had different disparities, the stimulus appeared as a group of slanted planes similar to the tiles on a “roof”. This construction ensured that the virtual plane that passed through the center of the rectangles had the same disparity gradient in both configurations.

Our results are inconsistent with a simple depth averaging. If this were the case, then the four rectangles would look the same in the two configurations, and thus the global slant of the plane passing through these rectangles would be equally discriminable. Our results clearly showed a difference between the two configurations. Our interpretation is that the locally slanted rectangles of the “roof” configuration creates a large uncertainty at the edges of the rectangles, and this makes the global slant judgment difficult. In contrast, the locally fronto-parallel rectangles of the “staircase” configuration are well-defined in depth, and this makes the global slant judgment easy.

These results, together with the ones obtained in the previous experiments, lead us to believe that it is important to take into account the whole uncertainty distribution (not just its mean) of the estimated depth of each feature and of the objects built from these features. We

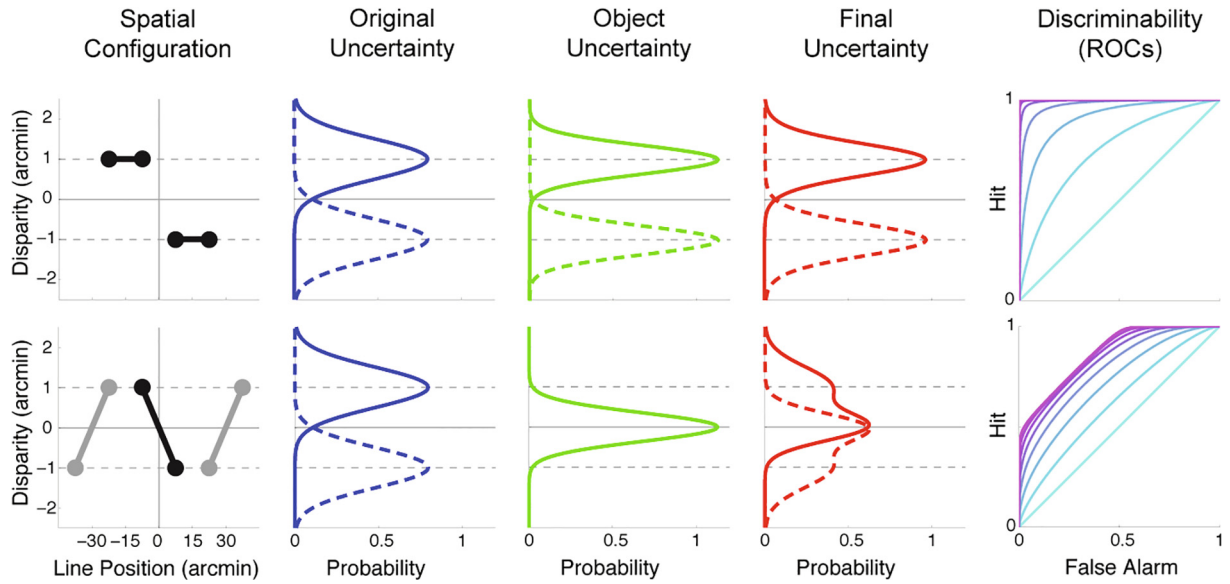


Fig. 13. Model of the influence of object formation on disparity uncertainty. Top row: the individual features whose depth is to be discriminated belong to two fronto-parallel planes. Bottom row: the same features are connected so as to form a slanted rectangle. We neglect here the contribution of the two other slanted objects on either side of the central slanted rectangle. For each row, the left-most column represents a plan view of the spatial configuration of the display. The following columns from left to right show the four stages of the model. First, the disparity uncertainty of each feature is estimated (solid blue line for left feature, dashed blue line for right). Then, a global object uncertainty (green curves) is constructed from its features' uncertainty. Here we assumed that the object uncertainty results from an optimal combination of all the features constituting the object. This object uncertainty propagates back to the individual feature in such a way that the final uncertainty (red curves) is a linear combination of the original and object uncertainties. The last column shows the ability to discriminate which feature is in front from their final uncertainties. Colored lines of the ROC plot show simulations obtained with different feature disparities (zero disparity in cyan and increasing disparities for more purple colors). In the bottom plot, the ROC curves are limited by an upper bound that causes a performance plateau (see Fig. 5A). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

propose a model based on this uncertainty principle in the next section.

7. Object uncertainty model

We develop here a model that can account for the detrimental effect of object formation on stereoacuity. In simple terms, the model relies on the idea that a global depth is computed for the whole object and this depth estimate then propagates back to the individual features of the object. In general, this procedure can help reduce depth uncertainty resulting from partial occlusion or noise corruption. However, when an object extends over several depth planes, the estimated depth of its front and back edges regresses to the global depth and their uncertainty increases.

The model can be conceptualized in four stages (Fig. 13). We describe the model in the context of the fronto-parallel and slanted-rectangle configurations of Experiment 1A. We then discuss how one parameter of the model can be used to characterize the strength of the grouping cue in the different experiments.

7.1. The four stages of the model

In the first stage, binocular disparities are estimated locally for each visual feature. Disparity uncertainty can be for instance represented within the population of binocular sensitive cells present in the primary visual cortex of cats and primates (DeAngelis, Ohzawa, & Freeman, 1991; Cumming & DeAngelis, 2001). We model this disparity uncertainty as a Gaussian probability centered on the displayed disparity and with standard deviation σ . For instance, if we are considering two lines to the left and right of fixation, with absolute disparities d_0 either in front (F) and behind (B) fixation, the probability density function (PDF) of their disparity uncertainties are

$$\begin{cases} \varphi_F = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d-d_0)^2}{2\sigma^2}\right) \\ \varphi_B = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d+d_0)^2}{2\sigma^2}\right) \end{cases} \quad (1)$$

In the second stage, a global depth is estimated for each grouped object. A strong assumption of the model is that the scene has been somehow segmented into objects. We are oblivious about the way this segmentation occurs, but we are aware that segmentation is a difficult problem in itself. In the line drawings used here though, the objects are defined by closure and are easy to segment. We hypothesize that this global depth is obtained by optimal combination of the individual feature uncertainties. Assuming independent disparity estimates for each feature, the object uncertainty is simply the product of each feature uncertainties. Because we assumed that the variance was the same for the two features, the uncertainty of the object constructed from these two features will follow a Gaussian PDF whose variance is halved (e.g. Ernst & Banks, 2002). We then need to consider two cases, depending on whether the two features belong to a fronto-parallel object (same disparities) or to a slanted object (opposite disparities). If they have the same disparity, for instance the front disparity, the PDF of the disparity uncertainty of this object that spans features “within” a fronto-parallel plane is

$$\varphi_{within} = \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{(d-d_0)^2}{\sigma^2}\right) \quad (2)$$

If instead the two features have opposite disparities ($+d_0$ and $-d_0$), the PDF of the disparity uncertainty of this object that spans features “across” two distinct depths is

$$\varphi_{across} = \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{d^2}{\sigma^2}\right) \quad (3)$$

In the third stage, the global depth propagates back to the elements forming the object. There are numerous ways this propagation could be

implemented, and here we shall simply use a linear combination (weighted sum) of the global disparity uncertainty and the original local disparity uncertainty. This linear combination is characterized by a weight γ of the influence of the object depth on the final disparity uncertainty of the local feature. For instance, for a slanted object that spans features across two depths, the new PDFs of the uncertainty of the front and back features are

$$\begin{cases} \varphi'_F = \gamma\varphi_{\text{across}} + (1 - \gamma)\varphi_F \\ \varphi'_B = \gamma\varphi_{\text{across}} + (1 - \gamma)\varphi_B \end{cases} \quad (4)$$

This parameter γ can be seen as a *grouping saliency index*: the stronger the grouping of the features, the larger the influence of the object on its individual features. As a side note, because the final depth uncertainty is a linear combination of the original and object uncertainties, the resulting final uncertainty may end up having two modes. It would be interesting to test whether this bimodality generates bistability similar to what can be observed when there is a large conflict between binocular disparity and other monocular depth cues (van Ee, Adams, & Mamassian, 2003).

In the fourth and last stage, a decision is made about which feature is in front. In this final stage, we compute the probability that the feature to the right of the fixation is in front of the feature to its left. Intuitively, we look at the certainty that the PDF of the right feature is more negative than the PDF of the left feature. We can compute this probability by estimating the so-called Receiver Operating Characteristic (ROC; Green & Swets, 1966). For instance, in the illustration of Fig. 13, the feature on the left was physically in front. We thus calculate the Hit probability as the cumulative distribution function (CDF) of the function φ'_F (solid red curve in Fig. 13), and the False Alarm probability as the cumulative distribution function (CDF) of the function φ'_B (dashed red curve). The probability to decide that the left line is in front is then the area under the ROC curve constructed from these Hit and False Alarm probabilities.

7.2. Fitting of the model to the experimental data

The model was fitted to the data collected in the psychophysical experiments. Given that the first parameter σ represents the disparity uncertainty of a line, it was adjusted on its own to the psychometric function from the baseline configuration consisting of two lines presented in isolation. The second parameter γ was estimated from the best fit of the model to the data of the other two configurations. This parameter was estimated to be 0.663 (95% confidence interval: [0.611, 0.713]), indicating that more than half of the final depth uncertainty of elementary features is corrupted by the object they belong to. The best fits are shown as solid lines in the results plots of Experiment 1A (Fig. 5A).

The model accounts for the main aspects of our psychophysical experiment. When the two lines on which the depth judgment is performed belong to two separate fronto-parallel rectangles, the object disparity uncertainty has a smaller variance thanks to the rule of optimal combination of the likelihoods. As a result, the individual features benefit from the less uncertain disparity of the object they belong to, and sensitivity to discriminate the depth of the two central lines increases. Instead, when the two lines belong to a single slanted object, the object uncertainty is centered on the mean disparity of the two lines, i.e. zero. The two individual features are then corrupted by this estimate of the object disparity, and sensitivity decreases. Importantly, there is a fraction of the final feature uncertainties (the red curves in Fig. 13) that is common to both features (represented by the parameter γ). Because this fraction is common to both features, it cannot contribute to discriminate their depth. As a result, performance will never be able to reach ceiling level no matter how large the lines' disparities are.

7.3. The origin of the depth bias due to object formation

Our model explains sensitivity changes to discriminate depth as a propagation of disparity uncertainty of the object to its individual features. Another property of the model is a bias of the estimated depth of the individual features introduced by the global depth of the object. This prediction was tested with the same participants in Experiment 1B where they viewed two parallel slanted planes (Fig. 4B).

To get an intuition for the prediction of the model, let us consider the condition where the planes are slanted to the right (with their right edge far). When the central vertical lines whose depth is to be discriminated are placed at zero disparity, participants should be at chance performance, but the model predicts that the left object will pull the left vertical line in front and the right object will push the right line to the back. A physical disparity has to be introduced on the two objects to cancel this bias and thus the psychometric function should present a shifted point of subjective equality (PSE). The model predicts that as slant increases, this PSE should increase. Fig. 5B shows the results of this experiment for five slant conditions, together with the predictions of the model with the parameter values set to those inferred in Experiment 1A (Fig. 5A). Without any free parameter, the model explains nicely both the changes in precision (sensitivity) and accuracy (bias) across the five slant conditions.

7.4. Grouping saliency indices for different grouping cues

Different cues can indicate more or less convincingly that two features belong to the same object (Wagemans et al., 2012a). We can determine the strength of a grouping cue thanks to our grouping saliency index in our object uncertainty model. When this index is zero, the object has no influence whatsoever on the individual features; when its value is one, the features lose their identity and instead behave as the whole object. We fitted the model to the different experiments, and extracted the grouping saliency index in each condition (Fig. 14).

In Experiment 2A, grouping was achieved with horizontal segments. We applied our two-parameter model to the psychometric functions in order to extract the grouping saliency index. This index reached on average 0.842, a value a bit larger than the one found in the first experiment (0.663) that might be explained by the fact that in this experiment there were multiple objects presented side by side. The grouping saliency index did not change significantly across the range of object sizes.

In Experiment 2B, grouping was achieved by color similarity. The grouping saliency index reached now on average 0.623, a value only a bit smaller than when horizontal lines were used to group features. Color similarity is therefore a strong grouping cue that leads to large deficits in perceived depth of the individual features forming the grouped object.

In Experiment 3, compatible configurations were achieved by pairing the last two lines of a temporal sequence with different pitch, and incompatible configurations paired the last two lines with the same pitch. The grouping saliency index reached on average 0.172, a value smaller than that obtained with the visual cues, but still larger than zero. In other words, auditory cues are sufficient to group lines together and impair the perceived depth of these lines.

In Experiment 4, observers had to discriminate between contraction and expansion of two lines in a two-frame motion. There were four different configurations. A cumulative Gaussian was fitted to the baseline “2-lines” configuration so as to extract the sensory noise that determines the sensitivity to discriminate expansion from contraction motion. This sensory noise was then used, together with a single grouping saliency index parameter, to fit the “2-rectangles” and “3-rectangles” configurations. Best fits are shown in Fig. 10. The smooth curve in the “1-rectangle” configuration is just a duplicate of the one from the “3-rectangles” configuration. The grouping saliency index reached on average 0.203, a value similar to the one obtained for cross-

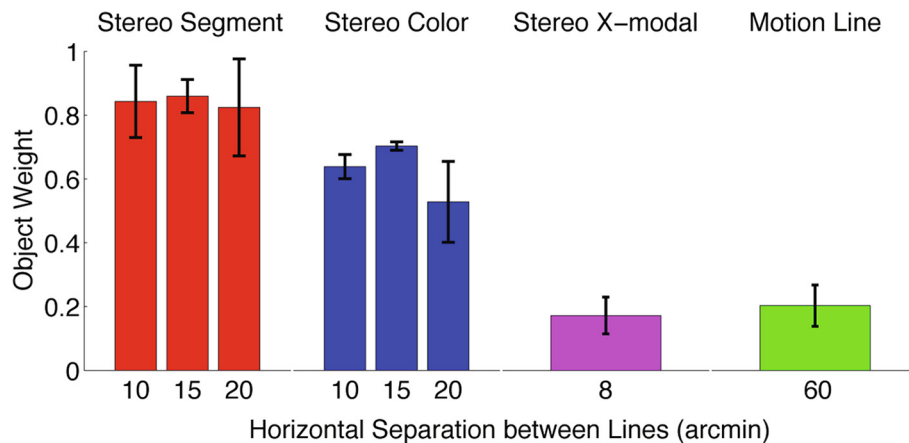


Fig. 14. Grouping saliency indices (parameter γ) for various cues. The stereo segment stimulus (red histogram) is the one used in Experiment 2A where vertical lines were connected by small horizontal segments. The stereo color display (blue histogram) is the one used in Experiment 2B where lines were grouped by color similarity. For these two experiments, three separations were used between the vertical lines. The stereo cross-modal experiment (purple bar) is Experiment 3 where lines were presented sequentially and paired with tones of compatible or incompatible pitches. Finally, the motion line experiment (green bar) is Experiment 4 where the task is based on lateral motions rather than binocular disparities. All experiments produced significant effects of object formation on the precision to locate the individual features of the object. Error bars are 95% confidence intervals on the estimated parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

modal grouping, and still larger than zero. Therefore, the motion result is similar to the one obtained in the stereopsis domain, although the effect is less striking.

7.5. Alternative interpretation of the object uncertainty model

The third stage of the model was described as a propagation of the object uncertainty to the individual features. An alternative interpretation, mathematically equivalent, is in terms of a mixture model (Knill, 2003; Körding et al., 2007; Orhan & Jacobs, 2013). In a mixture model applied to our task, two scenarios are entertained, one where the two features are processed independently and the other where the two features are combined into a single object. On each trial, there is a probability $(1 - \gamma)$ that the depth judgment is based on the individual features and a probability γ that it is based on the common object. In the latter case, because the depth uncertainty distribution is the same and centered on zero in the slanted condition, performance is at chance for these trials. This explains the plateau performance observed in the slanted condition.

8. Alternative explanations

The propagation of object uncertainty is sometimes detrimental and thus might appear counter-intuitive. Therefore, it is worth considering a number of alternative explanations to our model. The first of these alternative explanations is coming from an apparent cue conflict between binocular disparity and linear perspective. We then discuss two alternative explanations and two more models that rely on imposing some prior constraint on depth or surface orientation.

8.1. Linear perspective cue conflict

A natural alternative explanation to consider is one based on cue conflict. In Experiment 1, we used horizontal segments to group the vertical lines into rectangles. When the rectangles are fronto-parallel, the connecting segments should indeed be horizontal. But when the rectangles are slanted in depth, the segments should be tilted because of the laws of linear perspective. Therefore, presenting horizontal segments for all stimuli created a depth cue conflict in the slanted configuration that should reduce performance in that condition. Isn't this a simpler explanation for the detrimental performance observed in the slanted configuration?

First, we should keep in mind that linear perspective information is mostly relevant for large slants and large objects (Gilliam & Ryan, 1992). When there is a large conflict between linear perspective and binocular disparity information, human observers experience bistability between two percepts that are either mostly driven by linear perspective or binocular disparity (van Ee et al., 2003). In our setup, a disparity of 1.22 arcmin (the obtained threshold for the baseline condition) corresponds to a slant of 50.3 deg. However, because of the small size of the objects (two consecutive lines are separated horizontally by 15 arcmin), this slant translates to a difference in line heights of only 0.041 mm (0.13 pixels) for a line height of 15.5 mm (i.e. a 0.26% change in vertical length; see Appendix). A similar reasoning was used to rule out vertical disparities as the source of the effects in the Westheimer-McKee phenomenon (Mitchison & Westheimer, 1984). Even though these vertical disparities are negligible, it was reported that adding them improved depth discrimination thresholds in the closed figure configuration, but even then, they were still much worse than in the two lines configuration (Zalevski, Henning, & Hill, 2007).

We note that in the original demonstrations of Westheimer (1979), McKee (1983), and Mitchison and Westheimer (1984), the length of the vertical lines was about the same as the distance between the lines, so that when the vertical lines were connected, they formed a square figure (Fig. 1). It has been argued that this square, interpreted as a monocular object, is by default interpreted as a fronto-parallel square rather than a slanted rectangle (Stevens & Brookes, 1988). This fronto-parallel interpretation made more difficult the depth order estimation of its left and right edges. Valid as this observation may be, our stimulus was much higher than wide, and thus such an explanation based on a cue conflict with the monocular figure does not apply here. Therefore, we conclude that linear perspective information is negligible here and cannot be the source of the changes of sensitivity across experimental conditions.

8.2. Depth averaging

Do our results simply reflect a regression of perceived depth to the mean depth of the object? In the words of McKee (1983), “continuous figures could constitute a powerful input to a ‘global’ fusion mechanism which might average the disparities of the component features to assign a single depth value to the figure as a whole” (p. 197). Our model predicts this regression to the mean thanks to the second stage that integrates the disparities of all the features of an object. However, our

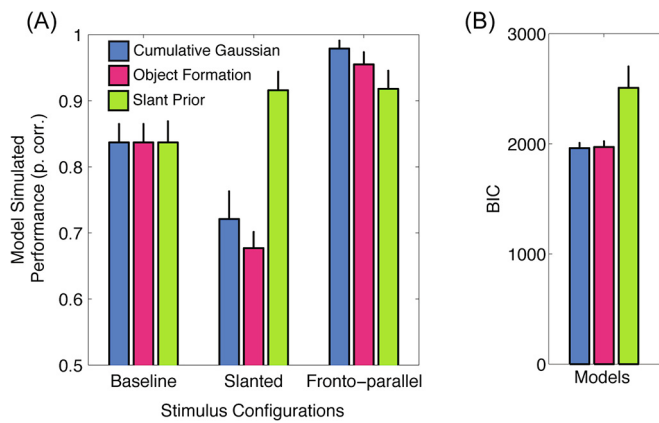


Fig. 15. Model comparison for the results of Experiment 1A. (A) Simulated performance for three models (different colors) for the three stimulus configurations (blocks of bars). Performance here refers to the probability correct of judging that one line is in front of another when these two have a physical disparity of 2 arcmin. The first bar (blue) shows a Cumulative Gaussian fit with 4 degrees of freedom, a standard deviation of the Gaussian for each of the three stimulus configurations and a lapse rate for the slanted configuration. The second bar (red) shows the performance of our object uncertainty model with 2 degrees of freedom. The third bar (green) shows the performance of the alternative model based on a priori for zero slant with 2 degrees of freedom. (B) Bayesian Information Criterion (BIC) for the three models in (A). Error bars are 95% confidence intervals based on 1000 bootstraps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model goes beyond this property in predicting also a larger uncertainty in the perceived depth of features belonging to slanted rather than fronto-parallel objects. This prediction comes from the third stage of the model where object uncertainty propagates back to the individual features.

The increased variability of depth estimates when features belong to slanted objects was observed in Experiment 1B. In that experiment, we found a shallower slope of the psychometric function when displayed slant increases (Fig. 5B). The increased variability was then directly tested with Experiment 5 where the averaged depth of the four local objects was identical in the two stimulus configurations, the “staircase” and the “roof” stimuli (Fig. 12B). In spite of being equivalent in terms of averaged depths, the global slant was much more difficult to estimate in the “roof” configuration because of the larger object uncertainty of the individual slanted planes. In conclusion, the effects of object formation cannot be reduced to depth averaging but need to include also some propagation of object uncertainty back to the elementary features.

8.3. A planar figure serves as a local reference plane

Mitchison and Westheimer (1984) found that depth discrimination was poor for a stimulus consisting in a square bracket where the end-points are within a single slanted plane (Fig. 1e), but good for a square bracket where the end-points are in fronto-parallel planes (Fig. 1f). This led the authors to argue that “measurements are made relative to a plane somehow defined by the visual system. [...] When a square is presented, we should argue that the plane is partly re-defined to be parallel to this square” (p. 1065). Careful measurements have confirmed that a slanted grid can indeed introduce some bias and change of sensitivity in the perceived depth of nearby objects (Glennester & McKee, 1999; Glennester, McKee, & Birch, 2002). However, we believe that this interpretation is incomplete to explain all of our results. For instance, in the slanted-rectangle configuration of our Experiment 1A, three slanted planes were presented with alternating slants ($-\sigma$ and $+\sigma$). Therefore, if one plane served as a reference plane, the slant of the next plane would be twice as large (2σ) and presumably easy to

detect thanks to a contrasting slant effect (van Ee et al., 1999). Yet this configuration was the most difficult to perceive.

8.4. Prior for fixation depth (zero disparity)

There are some reports in the stereo literature that perceived depth seems regressed to the fixation plane (e.g. van Ee et al., 2003; Zannoli & Mamassian, 2011; Hartle & Wilcox, 2016), as if observers had a prior for zero disparity. A model based on such a prior constraint does not include the concept of objects, so all three conditions of our Experiment 1A would be equivalent and produce identical sensitivities. Therefore, a model based on a prior for zero disparity cannot account for our results.

8.5. Prior for fronto-parallel objects (zero slant)

Instead of a prior for zero disparity, one may consider a model based on a prior for fronto-parallel objects. To implement this idea, we start with a representation of depth uncertainty similar to the one we used in the object uncertainty model. Namely, we assume that the depth likelihood of each vertical line is represented as a Gaussian function centered on the physical depth and with a standard deviation σ . When two lines separated laterally by $2L$ (the 2 comes the fact that the lines are symmetrically positioned relative to fixation) and in depth by $2D$ are connected, they form a slanted object (see again Appendix). The slant of this object is defined as

$$\text{slant} = \text{atan}(D/L) \quad (5)$$

The distribution of slant can thus be simulated by sampling the depth likelihoods of each line and applying Eq. (5). A prior for fronto-parallel object is represented by a Gaussian distribution centered on zero and with a standard deviation π . The posterior distribution of slant is then obtained by applying Bayes’ rule, namely by multiplying likelihood and prior and renormalizing so as to obtain a probability distribution function (Mamassian, Landy, & Maloney, 2002). From this posterior distribution of slants, we need to revert back to depth of individual elements. We can do this by inverting Eq. (5), or equivalently look for the depth uncertainties that could have generated the posterior distribution of slants.

We applied this procedure for both object configurations of Experiment 1A, the fronto-parallel and the slanted objects, and fitted the single π parameter (similarly to the main model, the σ parameter was first adjusted to the baseline condition). This model can indeed produce a reduced sensitivity for the slanted object, but it fails to account for our results for one fundamental reason. The model always generates full psychometric functions ranging from 0 to 1, and is incapable of producing plateaus similarly to what we obtained. In addition, the best fit of our data with this model is obtained with a π parameter close to 90 degrees that corresponds approximately to a flat (non-informative) prior over slants, so the prior does not even play the role of biasing slants towards fronto-parallel (Fig. 15). We thus conclude that a model based on a prior for fronto-parallel objects cannot account for our results.

9. General discussion

The process of object formation helps making sense of scattered visual information. However, once an object is formed, its individual features are perceived less accurately and less precisely. We have shown here that when an object is slanted, its front and back edges are both biased to the object global depth and more difficult to locate precisely. These accuracy and precision losses were evident for different stimulus configurations in stereopsis and also present in lateral motion judgments. The strength of these effects varied when different grouping cues were used, suggesting that the better the object is defined, the stronger is the contamination of the object on the individual features. In some cases, the loss of sensitivity could reach two orders of magnitude, a

dramatic effect even though the binocular disparity information presented to the observers remained identical across stimulus configurations.

9.1. Object grouping in line stereograms

The results of our experiments are consistent with a model that computes global properties of an object and then propagates the uncertainty distribution of these properties back to all the elements composing the object. In our experiments, the local and global properties are depth estimated from binocular disparities (Experiments 1, 2, 3, and 5) and lateral motion (Experiment 4). The model includes four stages. First the properties of local features are estimated with their mean and uncertainties. In the second stage, these estimates are combined within the object that the local features belong to. In the third stage, the global properties of the object propagate back to the local features and modify the uncertainty distribution of the local properties. More or less propagation is characterized by one parameter of the model, the grouping saliency index. Finally, a decision is reached based on the final uncertainty distribution of the local features.

We believe our model can account for the different manipulations of the Westheimer-McKee phenomenon illustrated in Fig. 1. Different depth discriminability thresholds can be obtained by adjusting the grouping saliency index in our model. This parameter represents different strengths of the grouping cues, and while this parameter was left free in our modelling, it would be interesting to estimate it in an independent experiment so as to constrain the model further.

Our model can also account for the results of Deas and Wilcox (2014) reviewed in the introduction, where the authors contrasted the two configurations shown in Fig. 3. Their closed object configuration is similar to our slanted-rectangle configuration in Experiment 1A, and so their decreased perceived depth can be explained by a regression towards the mean depth of the rectangle edges. In the extreme scenario where the grouping saliency index is 1, no depth would be perceived. Their segmented object configuration is more original as it compares a slanted object with a fronto-parallel one. In this case, the larger perceived depth would correspond to the depth between one edge of the fronto-parallel object and one edge of the slanted object. The former edge is perceived at its physical depth but the latter is regressed to the mean of the object it is attached to. In the extreme scenario where the grouping saliency index is 1, the perceived depth would be the distance between the fronto-parallel object and the mean of the slanted one. Therefore, the perceived depth of the segmented object configuration is always greater than that of the closed object configuration.

9.2. Object segmentation in RDS

We believe our model can also explain a number of other phenomena in the literature on random dot stereograms (RDS). The first phenomenon was described by Anstis, Howard, and Rogers (1978) who constructed a Craik-O'Brien Cornsweet illusion in depth rather than in luminance. They displayed a fronto-parallel plane that presented a sharp depth discontinuity in its center. The depth discontinuity was oriented vertically and was such that the left side came towards the observer and the right side went away. On the left side of the discontinuity, the surface was gradually receding backward to the depth of the fronto-parallel plane, and the right side gradually forward. When comparing the far left and right edges of the plane, the authors found a strong bias to perceive the left side closer than the right side. Interestingly, Anstis et al. (1978) originally argued that the same effect occurred when they rotated the surface by 90 degrees, thereby created a bias between the upper and the lower parts of the surface. However, this observation was later disproved by Rogers and Graham (1983) who found that the effect almost disappeared when the comparison was between the upper and lower parts of the display. These observations are important because they demonstrate large anisotropies between the

way the visual system processes horizontal and vertical gradients of disparity.

In a related study, Goutcher et al. (2018) showed to their observers RDS that displayed a gradual change of depth from the left to the right side of the surface. The change in depth followed a cumulative Gaussian profile whose standard deviation varied from trial to trial. Observers had to estimate the depth extent between the left-most and right-most edges of a test surface thanks to a standard stimulus that had a similar depth profile. Depth extent was overestimated relative to the standard when the test surface had a steep depth change, and inversely, depth extent was underestimated when the test had a shallow depth change. These results were explained by a hypercyclopean model that combines the responses of a set of neurons that have different disparity sensitivities over a range of spatial frequencies. A peak of sensitivity of 0.3 cycles/deg chosen from the literature gave a parameter-free model that matched the human data surprisingly well. However, the model broke down for their Experiment 5 based on a stimulus that was similar to that used by Anstis et al. (1978). This stimulus presented a depth discontinuity at its center, with gradual depth changes back to the sides of the surface. The authors found more depth underestimation as the depth discontinuity increased, while their hypercyclopean model predicted no depth bias.

We believe that our model can account for both the phenomenon described by Anstis et al. (1978) and the troubling stimulus of the study by Goutcher et al. (2018). The presence of a sharp depth discontinuity is a strong cue that the surface is composed of two distinct adjacent objects. Disparity averaging within each object then produces an overall object depth that is affected by the edge of the object on the side of the depth discontinuity. For instance, on the left side of the stimulus of Anstis et al. (1978) described above, the discontinuous edge coming forward creates an overall depth of the left object that appears closer to the observer, and reversely for the right side. This overall object depth when combined with estimated depth at the edges of the plane creates a bias that is consistent with the observed bias.

A depth discontinuity is a clear signal that there are two distinct objects on either side of the discontinuity. In binocular vision, these depth discontinuities often generate half-occlusions, and these parts of the scene that are visible by only one of the two eyes are strong cues for object segmentation (Anderson, 1994; Harris & Wilcox, 2009; Zannoli & Mamassian, 2011). However, even when there is no depth discontinuity, a smooth disparity gradient might be sufficiently good evidence for object segmentation if the gradient is steep enough. This issue was addressed by Cammack and Harris (2016) who displayed a square object that merged gradually with the fronto-parallel background. The gradient from the frontal square to the background could be more or less smooth, and increasing the smoothness of the gradient decreased the perceived depth of the square. The authors interpreted their results as evidence that observers were averaging binocular disparities over a very large area. Using model fitting, the best area over which disparities were averaged was of the same shape but slightly smaller (94.6% of original size) than the displayed square object.

Cammack and Harris (2016) seem to consider that the primary goal of object segmentation is to find regions of the visual field that have the same depth so as to be able to increase the signal to noise ratio by averaging over this region: “Ideally, averaging should be applied to regions likely to be of the same depth.” (p. 2). This is similar in spirit to the proposal of Mitchison (1988) who argued that the visual system was estimating planar surfaces for each segmented part of a visual scene. Instead, we consider that the goal of object segmentation is to find the proper boundaries of an object, and the depth averaging process that follows is there to provide a global depth estimate of the object in 3D space.

We believe that the model of Cammack and Harris (2016) misses the third stage that is present in our model, namely the propagation of object depth to the local depth estimates. Without this stage, any small depth variation would be ironed out. Therefore, if in their stimulus, the

central square did not have a constant depth but, say, contained a texture that followed a low amplitude corrugation in depth, their model would predict that the texture would disappear. In fact, while the value of the area over which averaging takes place is close to 100%, the difference might be evidence in favor of our model. Indeed, this difference leaves open the possibility that observers actually did segment the object properly with the correct size and averaged disparities over this whole area. The third stage of our model then takes this averaged disparity that is necessarily smaller than what the observers reported as peak depth, and combines it with the actual disparities in the center of the square object. The result of the combination can then match the observed results quantitatively thanks to an appropriate value of the object saliency index parameter of our model. For the thought experiment above with the corrugated texture, our model predicts an overall bias for smaller depth extent of the square and a smaller perceived amplitude of the texture, but the texture would still be visible (within the limits reported in the literature; Rogers & Graham, 1982).

9.3. Implementation of object depth propagation

The propagation of object depth uncertainty to elementary features leads to two important features that were present in our psychophysical results: (1) individual elements undergo a regression towards the mean, and (2) the uncertainty of individual elements grows when the object is composed of multiple dissonant parts. The first feature is about accuracy: the bias we found in Experiment 1B should be added to the list of binocular biases such as a preference for nearest disparity matching in the correspondence problem (Goutcher & Mamassian, 2005), a regression towards the zero disparity plane (e.g., Zannoli & Mamassian, 2011), or a bias resulting from an interaction with other cues consistent with zero depth (Held, Cooper, & Banks, 2012). The second feature is about precision: the large drop in sensitivity we found in Experiment 1A is characteristic of the Westheimer-McKee phenomenon and is spectacular when one appreciates that such changes of stereo sensitivity occur even though the same binocular disparity information is displayed.

The model we presented is only a first approximation to a recurrent network. In our model, depth uncertainty of individual features generates the global depth estimate of the object (second stage), and this global depth uncertainty propagates back to the features' uncertainty (third stage). But this new features' uncertainty should generate a new global object uncertainty that will propagate back to the feature uncertainty, and so on. Presumably this system will eventually reach a

stable state, and even though this dynamical analysis is important for a complete understanding of one's percept, it goes beyond the scope of the present study.

We have proposed that individual features lose part of their identity once the object these features belong to is formed. This idea is reminiscent of the principle of *explaining away* in Bayesian modeling, where two hypothetical objects compete to explain the presence of a feature in the image (Kersten, Mamassian, & Yuille, 2004). Once a feature has been assigned to an object, its presence is in a sense “explained”, and there is no need to pay more attention to it. It has been argued that the explaining away principle could be implemented as a top-down processing that attenuates the representation of explained elementary features (Murray, Kersten, Olshausen, Schrater, & Woods, 2002). Our model can be seen as implementing this attenuation in the propagation stage. Ideally, the global depth of the object is consistent with the depth of the individual features, and in this case the individual features are explained away by the object they belong to (no change of estimated depth occurs). If there is a modest discrepancy between the object and features disparities, this discrepancy is resolved by taking a weighted average of the estimates.

Instead of top-down processing, one may envisage that object formation could result from local processing within a dedicated cortical area. In that respect, cortical visual area V2 of primates appears to be a viable candidate since stereoscopic properties of some neurons in this area are strongly modulated by Gestalt principles (Qiu & von der Heydt, 2005). Whether the phenomena reported here correspond mostly to top-down mechanisms from higher cortical areas (Lamme & Roelfsema, 2000) or to lateral interactions within a cortical network is an interesting issue to be explored at the physiological level.

Acknowledgments

The authors wish to thank Giacomo Aldegheri, Marty Banks, Andrew Glennerster, Anna Ptukha, and Laurie Wilcox for comments on earlier versions of this work. This work was supported by a Chaire d'Excellence from the French ministry of research awarded to PM and ANR/NIH grant “Probabilistic models of perceptual grouping and segmentation in natural vision” (ANR-19-NEUC-0003-01). The psychophysical parts of this work have been presented at meetings of the Vision Sciences Society (2008, 2009 and 2012), and the modelling part at the European Conference on Visual Perception (2009).

Appendix

Linear perspective cue

The purpose of this Appendix is to derive the tilt angle of a rectangle rotated about a vertical axis. The slanted rectangle is constructed from two vertical lines on either side of fixation, one in front and the other behind fixation. The two vertical lines are connected by two segments that are horizontal in 3D space but that will appear tilted from the vantage points of the right and left eyes (Fig. A1). Let us denote by E the half inter-ocular distance, V the distance between the eyes and the monitor, and ψ the half-vergence angle. These three entities are linked by the equation

$$\tan(\psi) = \frac{E}{V} \quad (A1)$$

In our setup, the optical distance between one eye and the fixation point was 890 mm, and a typical value for E is 30 mm, so V is 889 mm.

Let L be the half-separation between two vertical lines. In Experiment 1A, L is 7.5 arcmin of visual angle or 1.94 mm in the plane of the monitor. When the rectangle is rotated clockwise (when seen from above), its right edge moves forward by a depth D . The rotation of the rectangle also shrinks the horizontal size of its projection on the monitor by an amount M_R for the right eye's image (the subscript R stands for the right eye). At threshold to reach 75% correct, we had to present on average a displacement M_R on the monitor that was 0.074 mm. Horizontal displacement M_R and depth D are linked by the equation

$$\tan(\theta_R) = \frac{M_R}{D} = \frac{E - L}{V - D} \quad (A2)$$

From this equation, we estimate D to be 2.34 mm. Knowing the depth allows us to estimate the slant σ of the rectangular object

$$\tan(\sigma) = \frac{D}{L} \quad (A3)$$

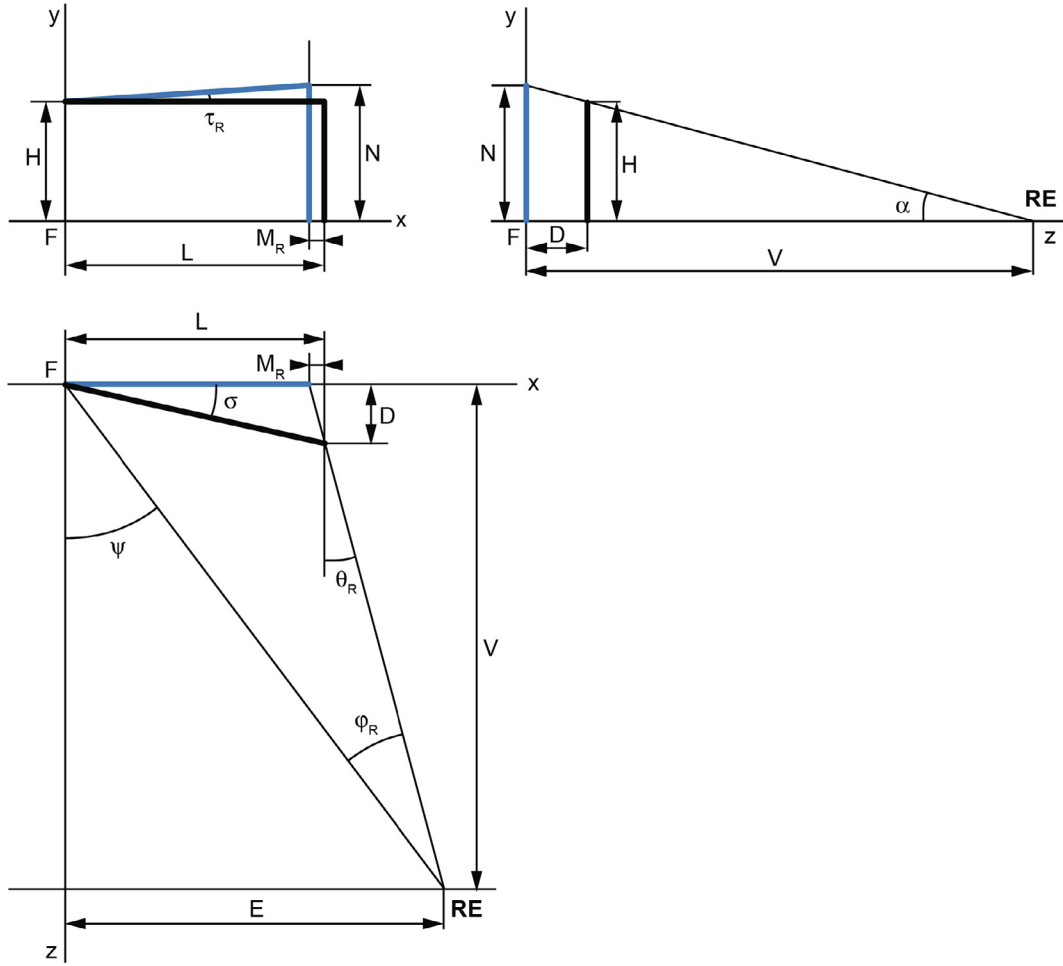


Fig. A1. Scene geometry for the projection of a slanted rectangle. The rectangle is composed of two vertical lines on either side of the fixation point (F), with the right edge in front and the left edge in the back. Only the top-right corner of the slanted rectangle is shown (illustration not to scale). The top-left panel shows the front view (using orthographic projection), the right panel the view from the left, and the bottom panel the view from above. The plane of the monitor onto which the images for the right eye and the left eye are displayed is the fronto-parallel plane passing through the fixation (only the geometry from the right eye (RE) is shown). The slanted rectangle is shown as thick black lines, and its projection onto the monitor is shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The slant of the rectangle at threshold is thus 50.3 deg.

From the angles ψ and θ_R , we can compute the angle φ_R subtended by the right-hand side of the rectangular object as seen from the right eye. We measure this angle from the fixation to the right edge of the object, with the convention that positive angles run counter-clockwise, so

$$\varphi_R = \left(\frac{\pi}{2} - \psi \right) - \left(\frac{\pi}{2} - \theta_R \right) = \theta_R - \psi \quad (\text{A4})$$

Likewise, the angle φ_L subtended by the right-hand side of the rectangular object as seen from the left eye is

$$\varphi_L = \left(\frac{\pi}{2} - \theta_L \right) - \left(\frac{\pi}{2} - \psi \right) = \psi - \theta_L \quad (\text{A5})$$

We can now define the disparity δ of the right edge (with the convention that crossed disparities are positive)

$$\delta = \varphi_R - \varphi_L = \theta_R + \theta_L - 2\psi \quad (\text{A6})$$

Using the approximation for small angles, we obtain

$$\delta = \frac{E - L}{V - D} + \frac{E + L}{V - D} - 2\frac{E}{V} = \frac{2DE}{(V - D)V} \quad (\text{A7})$$

The disparity at threshold is thus 36.7 arcsec (this is half of 1.22 arcmin reported in Experiment 1A because we are considering only the absolute disparity of the right line for the present calculation, whereas we reported relative disparities between the right and left lines in the results section).

Let H be the half-height of the vertical lines. In Experiment 1A, H is 30 arcmin of visual angle or 7.77 mm. The vertical projection N of the half-height H is obtained from the similar triangles that have in common angle α

$$\frac{N}{H} = \frac{V}{V - D} \quad (\text{A8})$$

At threshold, the ratio N/H is 1.0026, a very small deviation from unity. We note in passing that the height of this vertical projection is the same for both eyes. These projected lines are seen from different viewpoints, and so they will subtend different visual angles in the two eyes, but for the small lines near fixation used here, these visual angles are nearly identical and the resulting vertical disparities are extremely small.

The tilt τ_R of the top segment seen from the right eye is obtained from (reusing Eqs. (A2) and (A8))

$$\tan(\tau_R) = \frac{N - H}{L - M} = \frac{DH}{LV - DE} \quad (\text{A9})$$

At threshold, the tilt of the top segment is 0.628deg, a very small angle. From the values of the ratio N/H and the tilt τ , we conclude that the linear perspective cue is negligible in our stimuli.

References

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15, 122–131.
- Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, 367(6461), 365–368.
- Anstis, S. M., Howard, I. P., & Rogers, B. J. (1978). A Craik-O'Brien-Cornsweet illusion for visual depth. *Vision Research*, 18(2), 213–217.
- Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24, 2077–2089.
- Blake, R., & Wilson, H. R. (2011). Binocular vision. *Vision Research*, 51(7), 754–770.
- Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, 16(7), 263–268.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Buhmann, J. M., Malik, J., & Perona, P. (1999). Image recognition: Visual grouping, recognition, and learning. *Proceedings of the National Academy of Sciences USA*, 96, 14203–14204.
- Cammack, P., & Harris, J. M. (2016). Depth perception in disparity-defined objects: Finding the balance between averaging and segregation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 371(1697), 1–11.
- Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515–578.
- Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, 24, 203–238.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352, 156–159.
- Deas, L. M., & Wilcox, L. M. (2014). Gestalt grouping via closure degrades suprathreshold depth percepts. *Journal of Vision*, 14(9), <https://doi.org/10.1167/14.9.14>.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, USA*, 103, 18014–18019.
- Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189–211. <https://doi.org/10.1080/19345747.2011.618213>.
- Gilliam, B., & Ryan, C. (1992). Perspective, orientation disparity, and anisotropy in stereoscopic slant perception. *Perception*, 21, 427–439.
- Glennerster, A., & McKee, S. P. (1999). Bias and sensitivity of stereo judgements in the presence of a slanted reference plane. *Vision Research*, 39(18), 3057–3069.
- Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence for surface-based processing of binocular disparity. *Current Biology*, 12(10), 825–828.
- Goutcher, R., Connolly, E., & Hibbard, P. B. (2018). Surface continuity and discontinuity bias the perception of stereoscopic depth. *Journal of Vision*, 18(12), 1–15. <https://doi.org/10.1167/18.12.13>.
- Goutcher, R., & Mamassian, P. (2005). Selective biasing of stereo correspondence in an ambiguous stereogram. *Vision Research*, 45, 469–483.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception and Psychophysics*, 41(2), 117–158.
- Haefner, R. M., & Cumming, B. G. (2008). Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron*, 57, 147–158.
- Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research*, 49(22), 2666–2685.
- Hartle, B., & Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research*, 125, 64–75.
- Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, 22, 426–431.
- Hibbard, P. B. (2007). A statistical model of binocular disparity. *Spatial Cognition*, 15, 149–165.
- Hou, F., Lu, H., Zhou, Y., & Liu, Z. (2006). Amodal completion impairs stereoacuity discrimination. *Vision Research*, 46(13), 2061–2068.
- Howard, I. P., & Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. Oxford: Oxford University Press.
- Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: The University of Chicago Press.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271–304.
- Kesten, H. (1958). Accelerated stochastic approximation. *The Annals of Mathematical Statistics*, 29(1), 41–59.
- Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, 43(7), 831–854.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2(9), e943. <https://doi.org/10.1371/journal.pone.0000943>.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–579.
- Liu, Z. L., Jacobs, D. W., & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research*, 39, 4244–4257.
- Mamassian, P. (2008). Depth, but not surface orientation, from binocular disparities. *Journal of Vision*, 8(6), 89 Abstract of Vision Sciences Society meeting in May 2008. doi: 10.1167/8.6.89.
- Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13–36). Cambridge, MA: MIT Press.
- Masson, G. S., & Perrinet, L. U. (2012). The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neuroscience and Biobehavioral Reviews*, 36(1), 1–25.
- McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, 23, 191–198.
- Mitchison, G. (1988). Planarity and segmentation in stereoscopic matching. *Perception*, 17(6), 753–782.
- Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research*, 24, 1063–1073.
- Moscattelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, 12(11), 1–17.
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P. R., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 99(23), 15164–15169.
- Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience*, 10(5), 360–372.
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, 77, 2879–2909.
- Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review*, 120(2), 297–328.
- Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, 8(5), 379–391.
- Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *The Journal of Neuroscience*, 32(45), 15728–15736.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, 7(1), 379–412.
- Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron*, 47, 155–166.
- Rogers, B. J., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22(2), 261–270.
- Rogers, B. J., & Graham, M. E. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, 221(4618), 1409–1411.
- Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, 28(3), 371–386.
- Tsai, J. J., & Victor, J. D. (2003). Reading a population code: A multi-scale neural model for representing binocular disparity. *Vision Research*, 43, 445–466.
- van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *Journal of the Optical Society of America A*, 20(7), 1398–1406.
- van Ee, R., Banks, M. S., & Backus, B. T. (1999). An analysis of binocular slant contrast. *Perception*, 28, 1121–1145.
- Verghese, P., & McKee, S. P. (2006). Motion grouping impairs speed discrimination. *Vision Research*, 46, 1540–1546.
- Verghese, P., & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, 381, 161–163.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217.
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138(6), 1218–1252.
- Welchman, A. E. (2016). The human brain in depth: How we see in 3D. *Annual Review of Vision Science*, 2, 345–376.

- Werner, H. (1937). Dynamics in binocular depth perception. *Psychological Monographs*, 49(2), i–127.
- Westheimer, G. (1979). The spatial sense of the eye. Proctor lecture. *Investigative Ophthalmology and Visual Science*, 18, 893–912.
- Zalevski, A. M., Henning, G. B., & Hill, N. J. (2007). Cue combination and the effect of horizontal disparity and perspective on stereoacuity. *Spatial Vision*, 20, 107–138.
- Zannoli, M. (2012). Organisation of Audio-Visual Three-Dimensional Space. Unpublished PhD dissertation from Université Paris Descartes.
- Zannoli, M., & Mamassian, P. (2011). The role of transparency in da Vinci stereopsis. *Vision Research*, 51, 2186–2197.