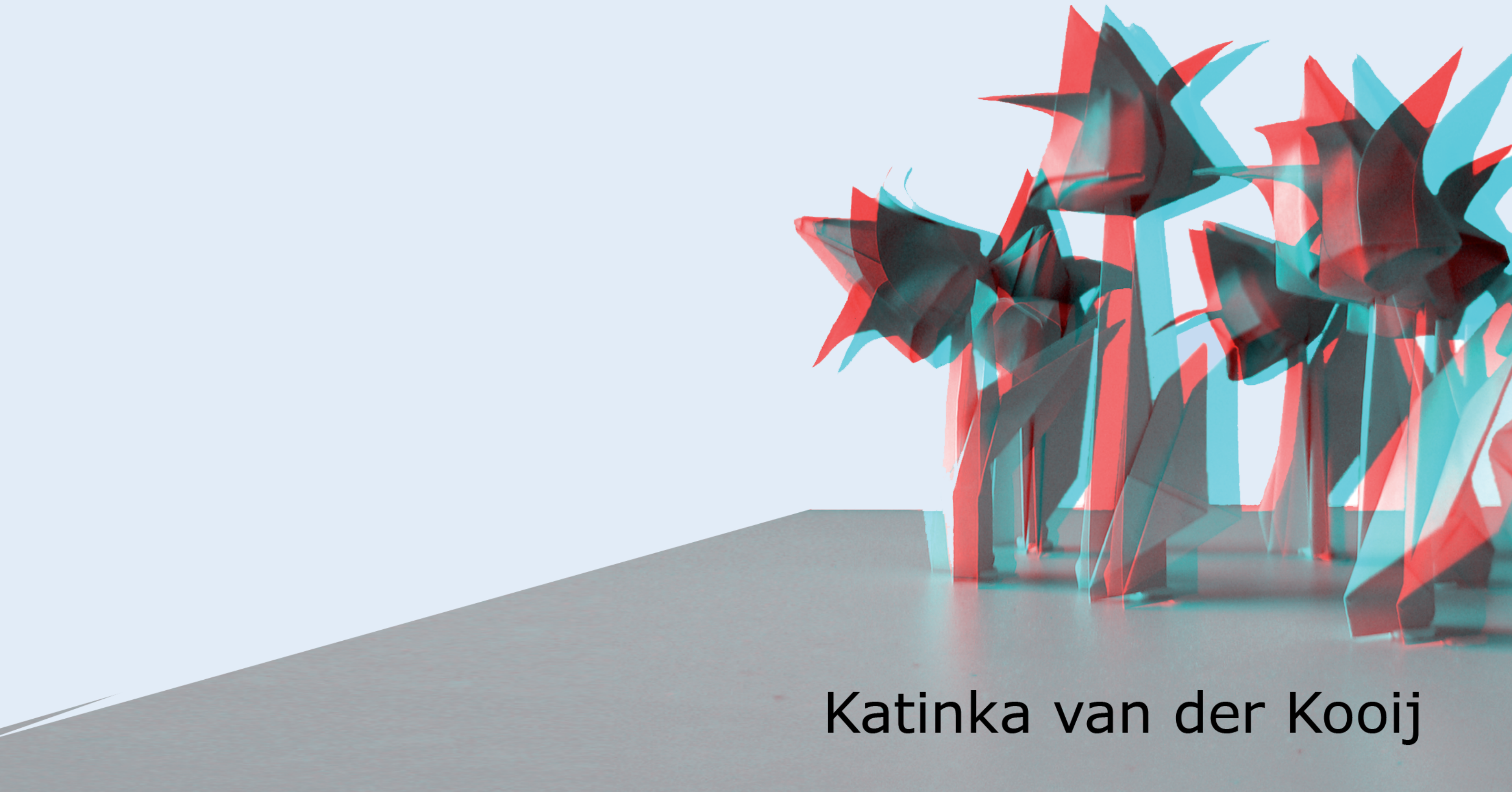
Perception of 3D Shape
in
Complex Scenes



Katinka van der Kooij

Perception of 3D Shape in Complex Scenes

Waarneming van 3D Vorm in Complexe Omgevingen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 13 oktober 2010 des middags te 2.30 uur

door Katinka van der Kooij

geboren op 4 augustus 1981
te Wageningen

Promotor: Prof.dr. F. A. J. Verstraten

Co-promotor: Dr. S. F. te Pas

The research in this thesis was financially supported by Grant 40004196 of the Netherlands Organization for Scientific Research (NWO) to Susan F. te Pas.

Table of contents

"Let us find other ingenious ways to transcribe objects onto flat surfaces; make the surface speak its own surface-language and not a false three-dimensional language which is alien to it."

Jean Dubuffet (1947)

1 General introduction

Visual perception seems to provide us direct access to the three-dimensional structure of our environment. We admire the curves in the face of someone we love, and caress them without the slightest concern that we will misplace our hand. Yet, scientific measurement of the relation between visual objects in the world and perception of these objects shows that veridical perception is more the exception than the rule. Visual perception is a mere interpretation of the physical world, although one that is astonishingly good for getting about and doing things. In demonstrations of visual illusions it becomes evident that perception may differ from the world truth. Many of such illusions are due to contextual effects in visual perception. An amusing example is the Ames Room, where the perspective of a room distorts the perceived size of the people in the room (Figure 1). In such contextual phenomena, perception shows itself a flexible phenomenon that adapts itself to environmental conditions.

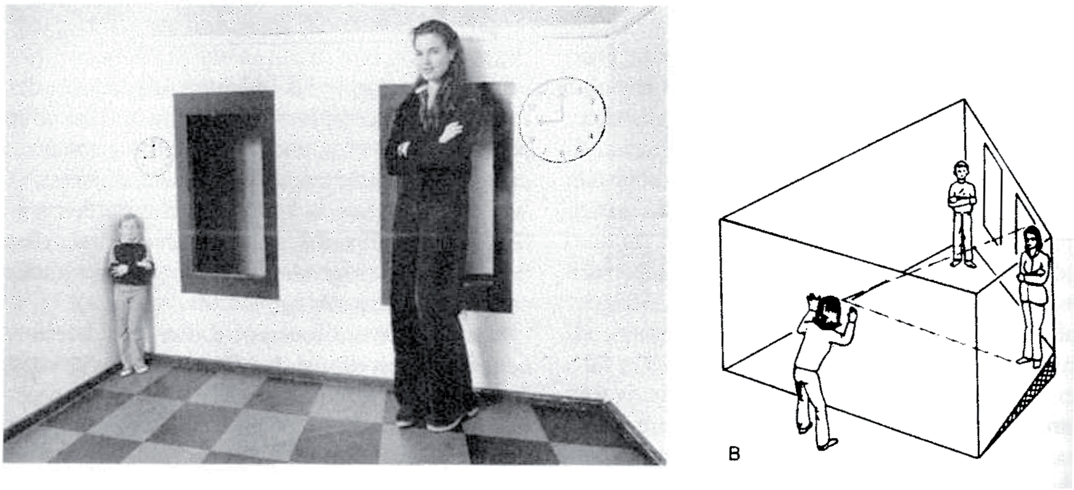


Figure 1. Ames room. A demonstration that misinterpretation of the room's perspective biases perceived size of the people in the room. Taken from Sekuler & Blake, 1998.

Contextual effects

Contextual effects are widespread in vision and can be broadly divided in two classes: contrast and assimilation biases. When information about a visual object is filled in with contextual information, this is called an assimilation bias. One of most well known phenomena of filling in is the 'watercolor illusion', where color information from contours fills in to the region encompassed by the contour (Pinna, Brelstaff, & Spillmann, 2001) Assimilation biases are widespread in visual perception and have also been found in the perception of motion (Ramachandran & Cavanagh, 1987) and orientation (Wenderoth & Johnstone, 1988). Especially when information about a visual object is unreliable, it makes sense for the visual system to fill in missing information with contextual stimuli. Points in space that are nearby often belong to the same visual object, and because information within an object is often correlated, contextual stimuli allow the observer to make an 'informed guess' about information at a central location (Schwartz, Hsu, & Dayan, 2007).

In a contrast bias, the difference between a visual object and its surroundings is perceptually enhanced. Contrast biases are most easily experienced in the color domain, and many artists have taken advantage of the color experiences that can be induced with surrounding colors. For instance, many French impressionist painters from the nineteenth century were known to have studied the rules of color contrast (Sekuler & Blake, 2002). Contrast biases are not confined to the color domain but occur in visual perception of features ranging from orientation (e.g. Wenderoth & Johnstone, 1988) to motion (see Reinhardt-Rutland, 1988 for a review) and shape (e.g. Rogers & Graham, 1982).

On the neural level, contextual effects have also been demonstrated. Already in 1968, Hubel and Wiesel showed that the response of a single neuron to a luminance stimulus could be modified by the stimulation of neighboring neurons. Typically, these interactions are of the inhibitory kind, where the neural response to a luminance patch will be less when neighboring neurons are also stimulated. Inhibitory interactions between neighboring neurons have been termed 'surround suppression' and have been investigated in many subsequent experiments in the cat and primate visual cortex. These experiments have shown that surround suppression occurs not only in neurons coding for luminance but also in neurons coding for more complex visual features such as orientation (Carandini, Heeger, & Movshon, 1997; Heeger, 1992; Ringach, 2010) and motion (Simoncelli & Heeger, 1998). Because responses to similar stimuli are suppressed, surround suppression is thought to make the visual system especially sensitive to differences between stimuli.

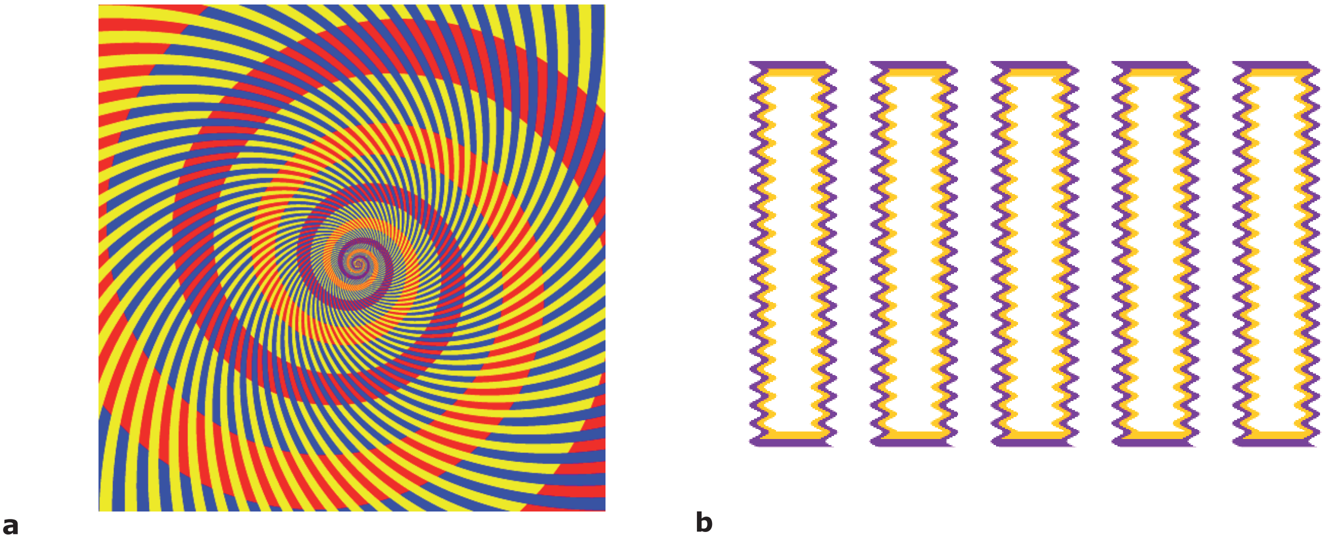


Figure 2. Color assimilation. **a)** The two red spirals appear of different hue, but they are the same color. **b)** Water color illusion. the inner section of the bar is filled-in with the yellow on the inner borders.

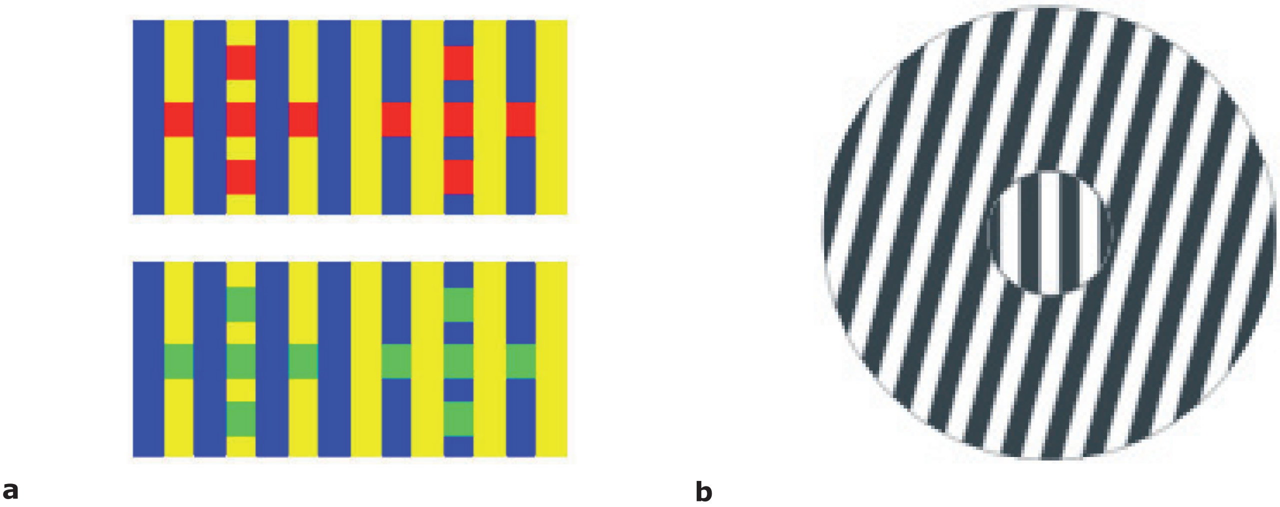


Figure 3. Contrast biases. **a)** Color contrast. The squares making up the cross appear of different color when they are embedded in the blue or yellow bars. **b)** Tilt contrast. The center grating is vertical, but appears tilted to the left relative to the rightward tilted surround.

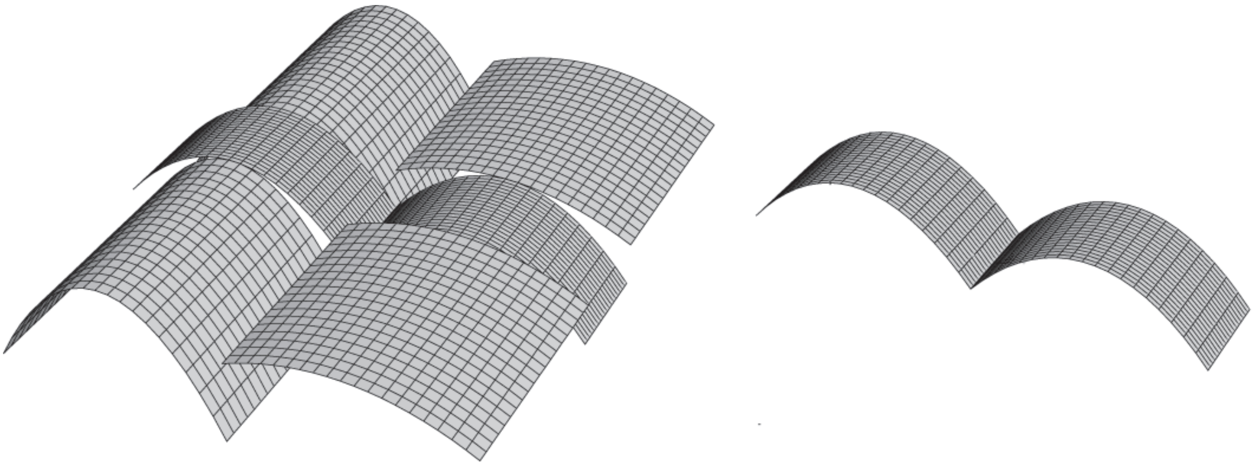


Figure 4. A contrast bias in curvature perception. The two sections of the middle strip are of equal curvature but the right section appears slightly more curved relative to its flankers.

The examples above show that contextual effects are a common phenomenon in visual perception. Furthermore, perceptual effects have strong neural counterparts which shows that contextual effects are firmly rooted in the cortical processing of visual information. At the start of this introduction we have raised the issue that despite the commonality of visual illusions, we normally have no problem acting on the three-dimensional environment. An intuitive solution to this problem is that, although perception of visual features such as color and luminance is biased, perception of the three-dimensional structure, within which these properties are perceived, is veridical. Yet, contextual biases also occur in the perception of three-dimensional space. Contrast biases in three-dimensional shape perception, where the shape difference with contextual stimuli is perceptually enhanced (Figure 5), have been measured in many psychophysical experiments and are surprisingly robust and similar between experiments (see Howard & Rogers, 2002 for review). Assimilation biases occur les frequently and are typically smaller in size as compared to contrast biases (Gillam, Blackburn, & Brooks, 2007; Poom, Olsson, & Borjesson, 2007; van der Kooij & Te Pas, 2009a, 2009b; van Ee, Banks, & Backus, 1999).

Even though contextual effects are widespread in visual perception, the standard approach to studying the mechanisms of shape perception has been to isolate an object from the environment and to study the relation between information in this object and mental phenomena. By isolating the object from its environment, the scientist can most easily determine the relation between information in this object and perception, allowing for precise hypotheses on the underlying mechanism. However, this approach deprives the visual system of contextual information, which might constitute a vital part of the information which it uses to perceive shape in natural scenes. The mechanisms of shape perception in complex scenes, where the visual system can use both direct and contextual information about shape, are the topic of this thesis. This topic is of general interest because it addresses the issue of how the visual system integrates information between spatially distant locations, one of the main open questions in vision research. A first step in finding the mechanisms of visual perception is to determine the processes by which a visual stimulus is translated into a visual percept. These processes are best determined with a psychophysical approach, where the relation between physical properties of world objects and mental phenomena is carefully examined. When both input (stimulus) and output (perception) of the visual system are known, information processing by the visual system can be described in mathematical models. Once we know how different types of information in a stimulus are related to visual perception of this stimulus, we can make generalized conclusions on how a specific type of information is processed.

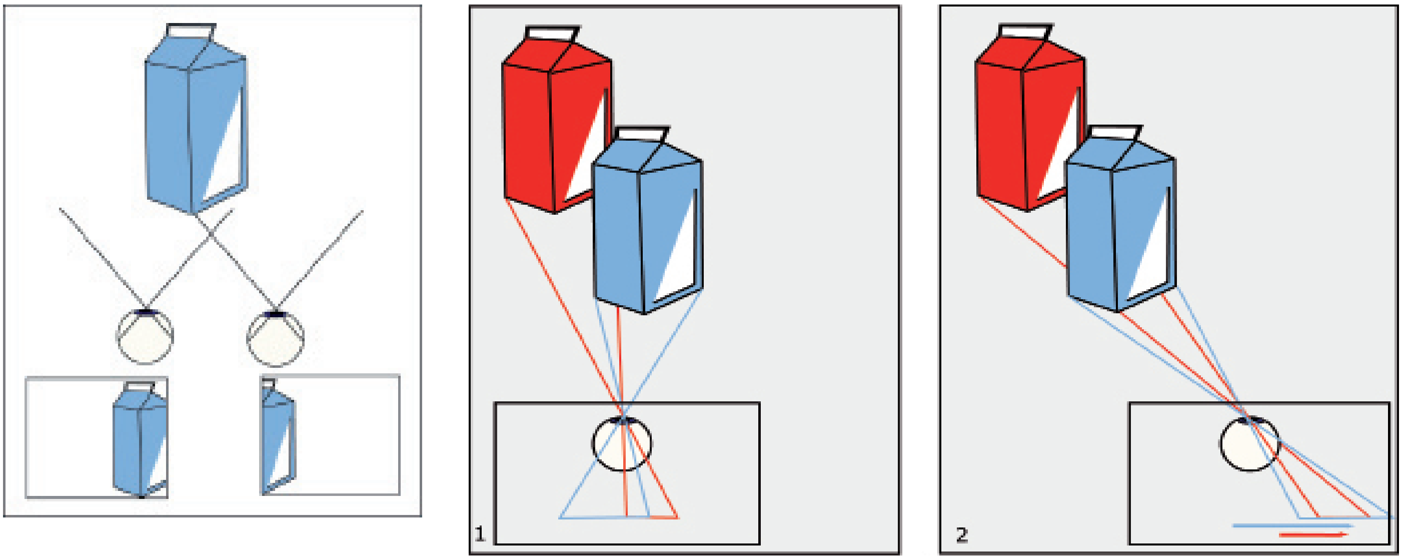


Figure 5.a) Retinal projections depend on the viewpoint of the two eyes. b) projected velocities depend on distance towards the observer.

Visual perception of three-dimensionality

An intuitive approach to determining the relation between the shape of world objects and visual perception would be to study the relation between perceived geometry and the three-dimensional geometry of these objects as they exist in the world. However, the visual system does not have direct access to the three-dimensional structure of the world but first has to sense visual information from projections onto the retina's. Because the three-dimensional stimulus is compressed onto a two-dimensional plane, these retinal projections are often dramatically different from their underlying world structure. Furthermore, a single three-dimensional object can cause range of two-dimensional projections and a single projection can belong to a range of three-dimensional structures. Because the mappings are one-to-many and many-to-one, it does not suffice for the visual system to learn the three-dimensional structure that a specific projections belong to.

The process of retrieving shape from retinal projections starts by reconstructing depth relations. The visual system can estimate depth from a wide range of 'depth cues', such as binocular disparity, motion, texture, perspective, shading and blur. Binocular disparity and motion are the best candidates to study how contextual information influences shape perception, because the processing of these depth cues is best understood. Detailed knowledge on how information in these cues is processed by the visual system allows us to precisely determine the effect that contextual stimuli have on the processing of this information. Binocular disparity relies on the fact that the two eyes view the world from a slightly different viewpoint (Figure 5.a). The spatial separation between the two eyes causes a shift in the retinal projections of an object in the two eyes. This shift depends on the distance to an object and thus gives provides a cue to depth. In 1959, Julesz showed that depth could be perceived from binocular disparity alone. He created a random texture of white dots on a black background (random dots stimulus), which contained no cues to depth in itself. But when he shifted the position of the dots between the left and right eye, observers got a clear depth percept (Figure 6). Since then, random dot stimuli have proved a potent tool to study the contribution of individual cues to depth perception. Motion of the retinal projections is another important cue to depth. When an observer moves through the environment, the projections of objects that are far away will translate faster over the retinas compared to the projections of objects that are nearby in space (Figure 5.b). Equally, when an object moves in front of a stationary observer, the projections of more distal parts of the object will translate faster over the retinas then the projections of nearer parts. Following this principle, depth from motion can be created in a random dot stimulus by translating dots in the texture at different velocities.

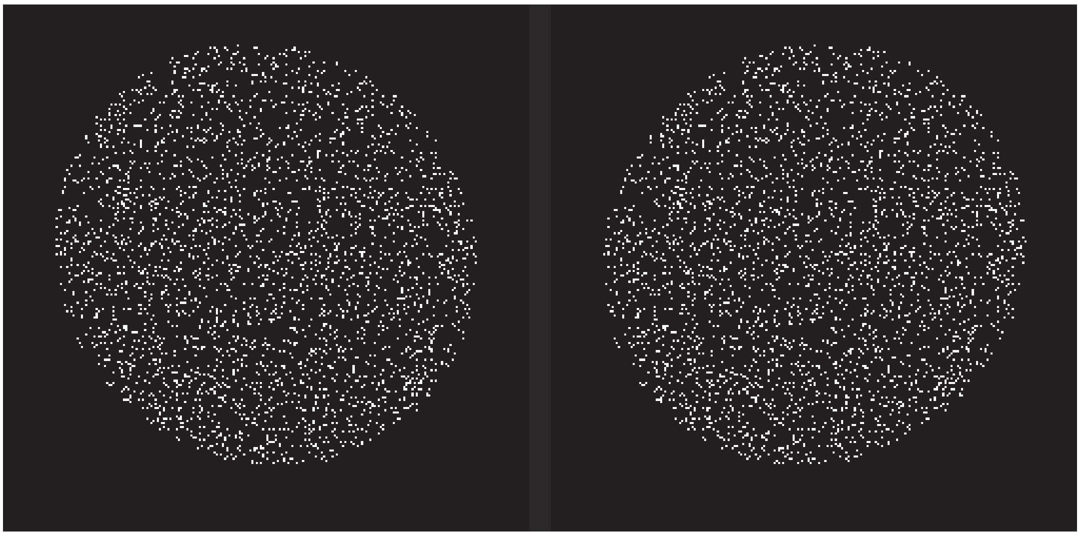


Figure 6. Random dot stereogram prepared for cross-fusion, depicting a surface slanted around the horizontal axis.

In the paragraph above we have seen how the visual system can retrieve depth relations from motion and disparity cues. These depth cues are 'sensed' by relatively early cortical areas. Sensitivity to binocular disparity has been found as early as in areas V3 and V3a (Neri, Bridge, & Heeger, 2004; Umeda, Tanabe, & Fujita, 2007). Temporal discontinuities are detected in V1 (Smith, Greenlee, Singh, Kraemer, & Hennig, 1998) but are elaborated into a motion signal with speed and direction in the medial temporal lobe (area MT/V5) (Beckers & Zeki, 1995). To recover surface shape, depth relations at different points in space have to be integrated. At the most basic level of integration, relative disparities and velocities are the difference in disparity or velocity between neighboring points in space. Next, the visual system computes gradients, which are a measure of how relative disparities and velocities change over space and specify surface slant. Neural sensitivity to these disparity and velocity defined gradients has been found in area MT (Nguyenkim & DeAngelis, 2003; Xiao, Marcar, Raiguel, & Orban, 1997) and the parietal cortex (Tsutsui, 2002).

Ultimately, a measure of surface shape can be retrieved by measuring a change in motion and disparity gradients (Howard & Rogers, 2002). Because motion and disparity each give incomplete information about shape, shape estimates from binocular disparity and motion may differ. Disparity, for instance, only specifies shape up to a scaling factor for viewing distance. To come to a unified shape percept, signals from different depth cues have to be integrated. If the visual system does not unify shape signals, this would result in an unstable shape percept, which switches between estimates based on individual depth cues. Although this situation can be induced in the laboratory (Knapen & van Ee, 2006), this is not what observers normally experience. If, on the other hand, the visual system deals with the inconsistencies between different depth cues by simply picking one and ignoring the other signals, it would not take full advantage of information in the environment. Empirical evidence shows that precision of shape perception from a combination of depth cues is higher then shape perception from any of the cues individually (e.g. Jacobs & Fine, 1999; Johnston, Cumming, & Landy, 1994; Vuong, Domini, & Caudek, 2006).

To conclude, shape perception consists of multiple processing stages. Information from individual depth cues has to be integrated over a spatial area and is combined with information from different depth cues. The basic problem that research on shape perception has dealt with, is that the visual system must somehow integrate information from these depth cues into a unified shape percept. Cue combination models describe how the visual system unifies shape signals from different depth cues into a coherent shape percept. A typical cue combination model approaches the problem of shape estimation from multiple depth cues as a statistical problem of estimating the source of multiple unreliable signals. In statistics, this is called a problem of 'maximum likelihood estimation'. In a maximum likelihood estimation model of shape perception, shape signals are averaged according to their reliability, resulting in a 'weighted average' of shape signals.

The stereogram in Figure 6 can be used to illustrate a maximum likelihood model of shape perception. In this stereogram, slant is specified uniquely by binocular disparity. At small distances, binocular disparity provides quite reliable information about depth and observers only slightly underestimate surface slant from binocular disparity, even though all other cues, such as texture and perspective, specify a flat surface. But when we step back and view the stereogram from a larger distance, the surface will appear less slanted. The decrease in slant percept occurs because binocular disparity becomes a less reliable cue and other cues signaling flatness, will gain influence. Maximum likelihood estimation models are popular in psychophysics, because the reliability of depth cues can be empirically estimated. When observers are given a shape discrimination task, where they have to indicate which shape is more curved, for instance, the smallest shape difference they can reliably discriminated is considered a measure of the reliability of the shape signal in the stimulus. When both the depth signals in the stimulus and the reliability of these signals are known, precise predictions on perception can be made. Furthermore, in the literature many examples can be found of maximum likelihood estimation models that adequately describe experimental data on perception of shape from a combination of depth cues (e.g. Curran & Johnston, 1994; Knill & Saunders, 2003; Landy, Maloney, Johnston, & Johnston, 1995). On the other hand, maximum likelihood estimation models only hold under the assumptions of normal distribution of shape signals and independence of variance between shape signals. As both assumptions are controversial, alternative models might describe the mechanisms of shape processing better.

First, implicit in the calculations of shape signal variance in standard maximum likelihood estimation models is the assumption that the visual system retrieves veridical shape. But not all shape signals are normally distributed compared to physical shape. Slant judgments from texture, for instance, are more reliable for steep slants compared to shallow slants. The slant difference in the two left images of Figure 7 is actually larger compared to the slant difference in the two right sections, but few observers perceive the differences this way. When plotted against physical slant, this causes a skewed distribution of the shape signal. Furthermore, the visual system might not aim to retrieve veridical shape as biases in shape perception are frequent, even in the perception of natural, full cue stimuli (Todd, 1995).

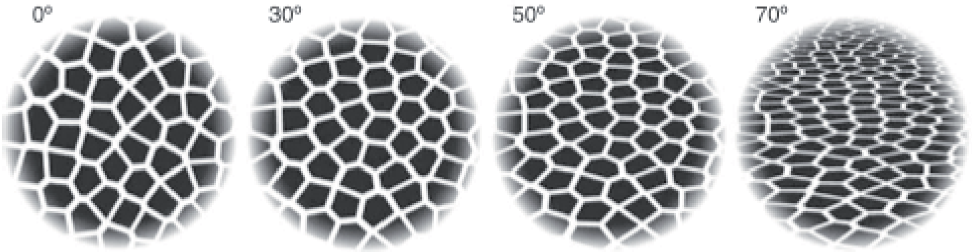


Figure 7. Slant from texture is more reliable for steep slants than shallow slant. Image taken from (Todd, Christensen, & Guckes, 2010)

Second, independence of variance between different shape signals is true only when these shape signals are processed independently. Because motion and disparity are computationally very similar cues (one depends on a difference in time whereas the other depends on a difference in space), the visual system might use a common structure to processes these depth cues. Indeed, neurons have been found in cortical area MT that respond to both disparity and velocity signals (Bradley & Andersen, 1998; Maunsell, 1983). Also, physiological measurements in this area have demonstrated both neurons that respond to disparity gradients (Nguyenkim & DeAngelis, 2003) and neurons that respond to velocity gradients (Xiao et al., 1997). Thus, the question whether motion and disparity are processed independently remains controversial (Bradshaw & Rogers, 1996; Domini, Caudek, & Tassinari, 2006; Landy et al., 1995). Hence, the mechanisms by which the visual system processes information from a combination of depth cues might be better described by a model that does not make the assumption of a maximum likelihood estimation model that motion and disparity cues are processed independently.

In the 'Internal Constraint' model of shape perception (Domini et al., 2006), motion and disparity are not processed independently. This means that the visual system directly makes an estimate of depth based on a combination of disparity and motion and does not first estimate depth-from-disparity and depth-from-motion before combining the two signals. Recent psychophysical studies (Di Luca, Domini, & Caudek, 2010; Domini & Caudek, 2009, 2010) show that this model is well able to predict human perception of depth from disparity and motion, including biases in perception.

To conclude, cue combination models have successfully described how the visual system combines information by different depth cues into a consistent whole, even though disparity and motion may be combined in different ways then originally thought. Current models of shape perception have in common that they do not incorporate contextual effects in shape perception. Yet, contrast and assimilation biases in shape perception show that contextual stimuli do have an effect on shape perception. In this thesis we would like to gain insight in the mechanisms that combine all contributing cues in a scene to a consistent whole. In the next paragraph, we outline the main open questions on contextual biases in shape perception that will be addressed in this thesis.

1.4 Mechanisms of shape perception in complex scenes

Contextual stimuli contain potentially valuable information about shape, such as relative information. Hence, one could envisage a cue combination model which combines not only shape information within a visual object but also relative information from contextual objects (van Ee et al., 1999). Because cue combination models have successfully described how shape in isolation is estimated from a combination of depth cues, we will start by testing whether such a model can also accommodate contextual effects in shape perception.

In a cue combination model of contextual effects in shape perception, relative (contextual) and direct information about shape could be combined at two levels of shape processing. Relative and direct information could be combined within the individual cue systems, which means that disparity information in contextual stimuli will only influence disparity information in a central stimulus, and not information by other depth cues (van Ee et al., 1999; Figure 8.a). Alternatively, relative and direct information about shape could be combined after combination of depth cues (Figure 8.b). In Chapter 2, we address the level of shape perception where

contextual information is integrated in the shape estimate. As explained above, in a basic random dot texture all cues specify a flat surface, but an impression of shape can be created by adding disparity or motion information. To test whether information by one depth cue is able to bias perception of shape from another depth cue, we create a stimulus where the shape of a central surface and flankers is defined either in the same or in a different depth cue and measure the contrast bias induced by the flankers. With this stimulus, we show that when a central surface and flankers are defined by the same depth cue (disparity or motion) a contrast bias occurs whereas an assimilation bias is found when the central surface and flankers are defined by a different depth cue. Now the level of shape perception where contextual effects take place has been determined, we proceed to investigate how the visual system combines direct and contextual information about shape.

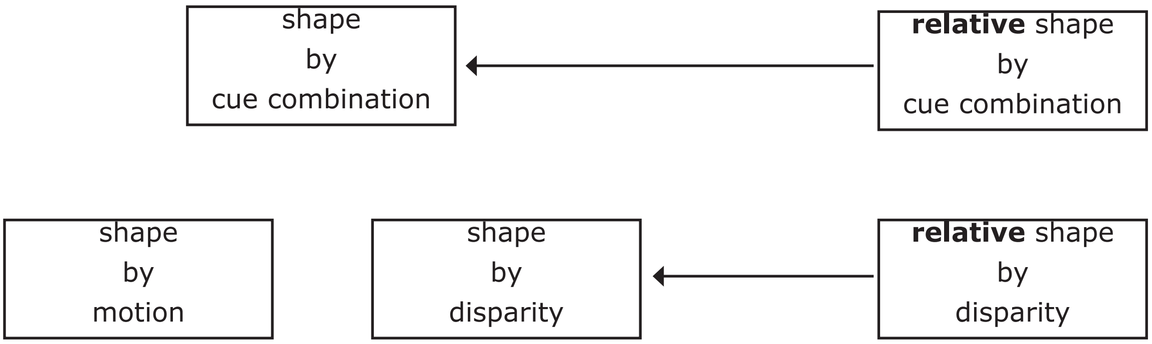


Figure 8. Two levels where contextual and direct shape could be combined in a maximum likelihood estimation model. **a)** contextual effects within cue systems **b)** contextual effects in shape from cue combination.

If cue combination takes the form of maximum likelihood estimation, shape signals are combined according to their reliability. This means that if contextual effects in shape perception are due to weighted combination of contextual and direct shape information, these effects depend on the reliability of shape signals. When contextual reliability is relatively low, contextual cues would have less effect on the shape estimate and, as a consequence, a smaller bias would be observed. In a random dot stimulus, the reliability of shape signals can be changed by adding spatial noise to the random dot pattern that defines a surface (Ernst & Banks, 2002). In Chapter 3, we change the reliability of contextual and central stimuli by adding shape noise to the surfaces. When the reliability of contextual stimuli is diminished with added noise, the contrast bias they invoke is indeed diminished. When the reliability of the central shape is diminished with added noise, however, an assimilation bias instead of a contrast bias is observed. The change in bias direction (contrast to assimilation) with lowered reliability of the central shape can not be explained by a maximum likelihood estimation mechanism of contextual effects in shape perception. Therefore we proceed to look for a different mechanism which better explains contextual biases in shape perception. To this end, we will study the relation between relative information in a stimulus and contextual biases in the perception of this stimulus.

The classical shape contrast stimulus consists of a central surface and flankers. Context in such a stimulus is limited to two adjacent stimuli, whereas context in the complex scenes of the natural environment contains much richer shape variation. To reveal the nature of visual mechanisms that produces contrast biases, it is important to determine whether only local shape differences affect the bias or whether the bias is affected by global properties of the stimulus. In Chapter 4, we address the spatial extent to which contextual information is integrated in the shape estimate. To this end, we present a central surface with not one, but five surfaces on each side. This creates a distribution of shape signals in the stimulus of which we vary both the mean and the variance. In such a stimulus, shape perception could be affected by exclusively the shape of adjacent stimuli or global shape properties of the stimulus. Chapter 4 shows that shape perception is affected by the global shape in the stimulus but not by the variation of shapes in the stimulus. As in Chapter 3, the finding that shape perception does not depend on variance in the stimulus is not consistent with an MLE model of shape perception. Instead, the mechanism that produces contextual effects in shape perception relies on relative differences with the global shape in a stimulus.

Contextual effects in shape perception may be caused by a mechanism that renders the visual especially sensitive to relative differences. On the neural level, sensitivity to relative differences is achieved by surround suppression where the neural response to similar stimuli is suppressed. Such surround suppression takes the character of divisive normalization where the common factor in a stimulus is averaged out and relative differences remain. This way, redundancy is a stimulus is reduced which allows for more efficient processing of the remaining information. A recent computational model (Schwartz, Sejnowski, & Dayan, 2009) shows how a mechanism of divisive normalization can produce contextual effects in tilt perception. In Chapter 5, we compare the predictions of an MLE model of contextual effects and the predictions of a divisive normalization model to empirical data on slant contrast biases reported in the literature. This analysis shows that contextual effects in slant perception are better explained by a mechanism that reduces redundancy by divisive normalization. In a final study we address an implication of this mechanism.

Meaningful redundancy reduction can be achieved by a mechanism of grouping and segmentation where information within groups is reduced, but where borders between groups are enhanced. If contextual biases are due to a mechanism of redundancy reduction, this suggests that contextual biases may depend on grouping dynamics in a visual stimulus. In Chapter 6, we address the issue of whether contextual effects in shape perception are related to the grouping dynamics in a stimulus. Because the depth separation between three-dimensional surfaces is already a strong grouping cue, grouping and shape cues maybe be confounded in a standard shape contrast bias. To avoid this confound, we study a contextual bias in the temporal domain. In an aftereffect, prolonged stimulation with an adaptation stimulus causes a contrast bias in a sequentially viewed stimulus. Between the adaptation and test stimulus we vary grouping by proximity and similarity. Observers adapt to a surface in depth and are tested with another surface of which we varied overlap in spatial position and surface contour with the adaptation stimulus. The adaptation stimulus caused a contrast bias in the perception of the sequentially viewed stimulus, but this bias did not depend on whether there was overlap in position or surface contour between the adaptation and test stimulus. This means that we find no evidence for a relation between grouping dynamics in a stimulus and contextual effects in the perception of this stimulus.

Chapter 2

Perception of 3D Shape in Context: Contrast and Assimilation

Published as: van der Kooij, K. & te Pas, S. F. (2009) Perception of 3D Shape in Context: Contrast and Assimilation. *Vision Research, 49*, p. 746-751.

Abstract

Whereas integration of shape and surround is held to occur through cue-dependent representations, we show that both cue-invariant and cue-dependent representations are involved. A central hinged plane and larger flanking plane were defined by either binocular disparity or motion. In a 'within-cue' condition, shape and surround were defined by the same cue and in a 'cross-cue condition' they were defined by a different cue. Observers compared the dihedral angle of the central shape with a constant reference. When the central shape was defined by disparity, the surround stimuli invoked a *contrast* bias in the within-cue condition, but shape *assimilation* occurred in the cross-cue condition. When the central shape was defined by motion there were overall no significant results, but if a contrast bias was observed, it was in the within-cue condition where integration could occur through cue-dependent representations.

Introduction

The basic problem of studying the perception of 3D shape in complex scenes is that visual context affects perceived shape. Context can help resolve shape ambiguity (Gilliam & Grove, 2004) and guide visual search (Brady & Chung, 2007) or improve shape perception (Todd, Thaler & Dijkstra, 2005). However, global aspects of a scene can also bias the perception of local three-dimensional object properties like depth, attitude and curvature. Clear examples of such influence are shape contrast effects, where the perception of shape contrasts with its surround. That is, it is biased in the direction opposite to neighbouring shapes. Such contrast effects can be found in a wide range of geometrical properties like depth, slant and curvature, demonstrating that is a general visual phenomenon (e.g. Gibson, 1933; Cornsweet, 1970; Anstis, 1975; Graham & Rogers, 1982). In this paper, we address the underlying dynamics responsible for the integration of shape and surround.

The brain has to infer 3D shape from information by different depth cues such as disparity, motion or texture and so-called cue combination models describe how the information from these cues is combined. The 'Modified Weak Fusion' (MWF) model by Landy et al. (1995) has ample support and states that shape estimates from different cue systems ('shape by texture' or 'shape by disparity' etc) are combined according to a weighted linear combination rule. This means that the model rejects interaction between cue modules before cue combination, as this would lead to nonlinearities in the combination rule. Yet, because shape estimates from different cue systems are qualitatively different, they cannot be meaningfully averaged and some interaction between cue systems must occur. In the MWF framework, such interaction would occur at a stage where missing parameters in one cue system are filled-in with parameters from another cue system. For example, the missing parameter of viewing distance in retinal disparity can be inferred using information from motion parallax (Landy et al., 1995). But there might be more fusion between cue modalities than implemented in the MWF model. Especially motion and binocular disparity are likely candidates for such strong fusion. On the physiological level, single cell recording in cat striate cortex has shown neurons responding to velocity disparity as well as spatial disparity (Anzai, Ohzawa & Freeman, 2001; Maunsell & van Essen, 1983). Qualitatively, motion and binocular disparity are also similar: motion can be inferred from a difference in time, whereas binocular disparity is based on a difference in retinal space. Consistent with such observations, an 'Internal Constraints' (IC) model (Domini et al., 2006) proposes that, in a first stage of shape processing, an estimate of three-dimensional structure is made by strong fusion of motion and binocular disparity signals.

As there is a debate on the degree of fusion between depth cues, several studies have assessed the cue-dependency of 3D shape representations. These studies provide support for the existence of cue-dependent (shape by cue x) as well as cue-invariant (shape by cue combination) representations. On the behavioural level, some report cross-cue adaptation of slanted surfaces (Bradshaw & Rogers, 1996), whereas others report that slant cues adapt independently (Knapen & van Ee, 2006). By comparing fMRI event-related adaptation effects with psychophysical behaviour, Welchman et al, (2005) showed that responses in retinotopic areas corresponded to changes in cue information whereas responses in extrastriate ventral and dorsal areas were related to changes in perceived 3D shape based on cue combination.

Here, we investigate the question whether cue-invariant as well as cue-dependent representations are involved in the integration of shape and surround. Illusions of shape contrast can be found in stimuli that are defined by different cues such as luminance, motion or disparity and are highly similar in these modalities (Curran & Johnston, 1996; te Pas & Kappers, 2001; te Pas, Rogers & Ledgeway, 2000). This at least suggests that these biases are independent of cue modality. If the mechanisms that integrate shape and surround indeed rely on cue-invariant representations, surround stimuli are able to invoke a bias in shape perception regardless whether they are defined by the same depth cue. But if shapes are integrated with their surround by cue-dependent representations, surround stimuli would not be able to invoke a bias in the perception of a central shape when they are defined by a different depth cue (the 'cross-cue' situation). The studies that have addressed such questions maintain that shapes are integrated with their surround by cue-dependent representations. Van Ee et al. (1999) propose a Slant Estimation Model, which explains slant contrast from weighted combination of a relative disparity cue and a direct shape cue according to the following rule:

$$S_t = w_{t,dir}S_{t,alone} + w_{t\_ind}(S_{i,alone} + S_{rel}).$$

The first term is a direct estimate that is based on signals created by the test shape alone and the second term is an indirect estimate based on signals created by the inducer alone and the relative disparity gradient between the inducer and the test shape ($w_{t,dir} + w_{t,ind} = 1$). This combination rule predicts contrast when shape and surround are both defined by disparity but assimilation when the inducer's disparity specified slant is zero whereas it's slant by monocular cues is nonzero. Consistently, van Ee et al. (1999) report a contrast bias when shape and surround are defined by disparity but assimilation when the inducer's slant is defined by pictorial cues. More recently, Poom et al. (2007) found slant contrast when shape and surround were defined by disparity but assimilation when the surround slant was defined by motion and pictorial cues. They explain these results in the light of the Slant Estimation Model, and claim that both slant contrast and assimilation are a by-product of weighted combination of a relative disparity-cue. But the contrast and assimilation biases described by van Ee et al. (1999) and Poom et al. (2007) could also be caused by a different mechanism: one relying on cue-dependent representations and the other on cue-invariant representations.

We hypothesize that there is a dissociation in the depth-cue dependency of contrast and assimilation of 3D shape properties where slant contrast reflects integration by cue-dependent representations whereas assimilation reflects integration by cue-invariant representations. To test these hypotheses, we present observers with a central test shape flanked by a larger surround shape. Test shape and surround could be defined by motion or disparity and were defined by the same (within-cue condition) or by a different depth cue (cross-cue condition). We compare the biases in shape judgments between cue conditions. This way, we are able to demonstrate dissociation in the cue-dependency of slant contrast and assimilation. First, if slant contrast occurs exclusively in the within-cue conditions, this is evidence that contrast is mediated by cue-dependent representations. Second, by including the novel condition where the central shape is defined by monocular cues and the surround by disparity, we can assess the Slant Estimation Model (van Ee et al., 1999), which attributes both contrast and assimilation to interactions between cue-dependent representations. The combination rule proposed by van Ee et al. makes different predictions for the two cross-cue conditions. When the inducer's slant is specified by monocular cues and the central shape by binocular disparity, it predicts assimilation, as found by van Ee et al. (1999) and Poom et al. (2007). But when the inducer's slant is specified by disparity whereas the central shape is defined by monocular cues it predicts *enhanced* slant contrast. If we find assimilation in this novel situation, this is evidence that assimilation is mediated by cue-invariant representations.

Methods

Stimuli

Stimuli were viewed with red/blue anaglyphs and depicted a hinged plane receding in depth, surrounded by a larger hinged plane, also receding in depth (Fig 1.). All shapes were 14.1 degrees visual angle high and were separated by a gap of 0.4 degrees visual angle. Central shape and inducer horizontally subtended 6 and 21.1 degrees respectively. All surfaces were projections of a rectangular random dot patch onto a hinged plane. Therefore, frontal stationary 2D images contained no texture cues to slant.

To evaluate the perceived shape of the test stimulus, observers made comparisons of a range (100 degrees to 140 degrees) with a constant reference stimulus with a dihedral angle of 120 degrees (method of constant stimuli). Observers judged, using a key-press, which of the two sequentially presented stimuli (test and reference order randomized) had a smaller dihedral angle. To avoid response strategies incorporating the surround, they were explicitly told that the inducers do not contain any information they need for their task.

The dihedral angle of the test and reference surround was either 90 degrees or 150 degrees, resulting in 4 surround conditions: two where the dihedral angle of the surround of test and reference shape was the same (90 degrees or 150 degrees), and two where it was different (with test and reference surround angle 90 degrees and 150 degrees or 150 degrees and 90 degrees respectively). As central shape and surround could be defined by binocular disparity or motion, there were 4 cue conditions: a ‘within-cue’ and ‘cross-cue’ condition for the case where the central shape was defined by disparity or motion. Shape-from-motion was created by simulating a surface rotation of approximately 16 degrees through 40 frames back and forth around a horizontal axis. Each frame remained on the screen for about 0.04 seconds.

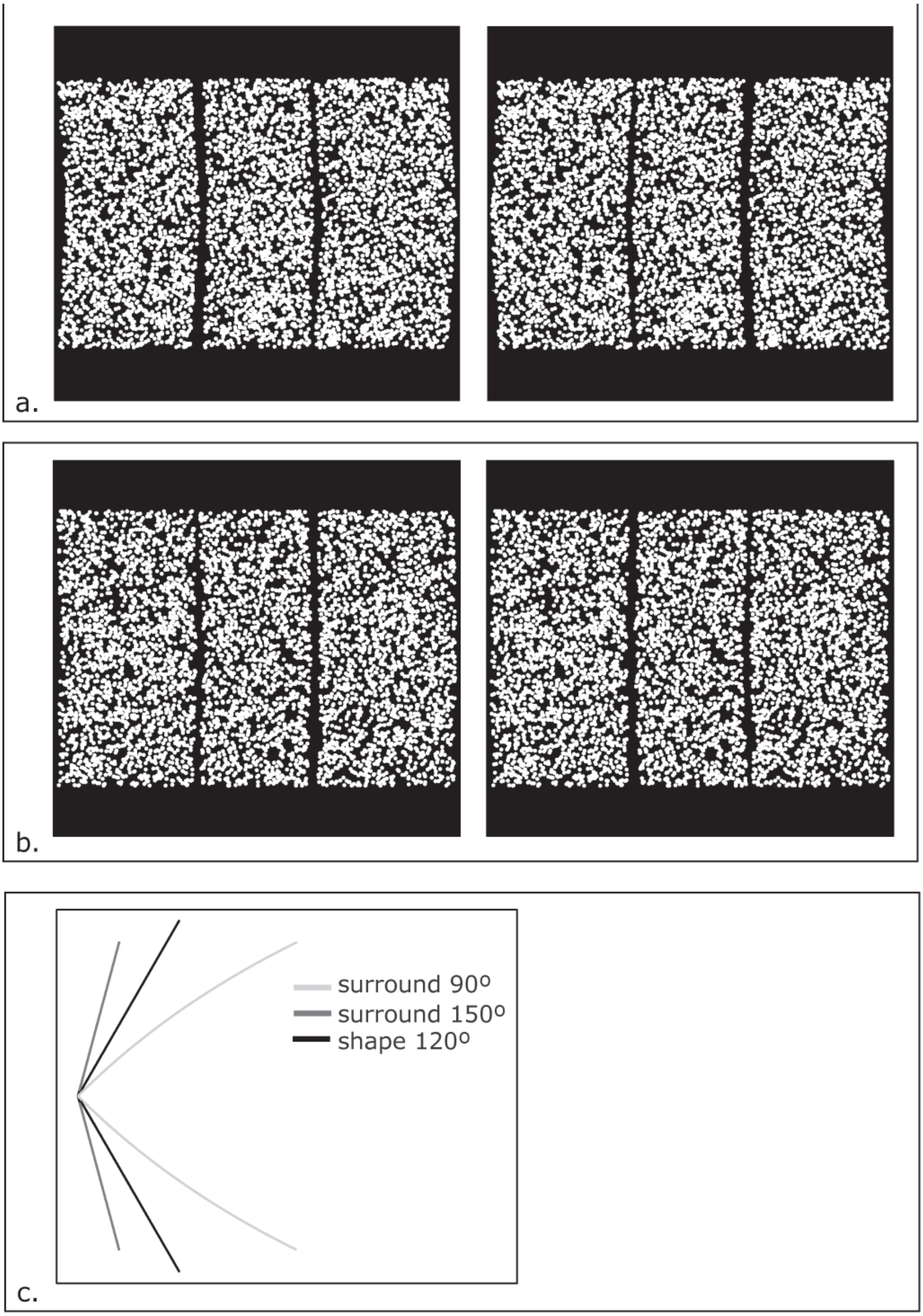


Figure 1. Stimulus set-up. **a.** stereogram of the situation where the surround dihedral angle 90 degrees. **b.** stereogram of the situation where the surround dihedral angle was 150 degrees. **c.** Side-view cartoon of the angle configuration. Anaglyph demonstrations of the stimuli can be viewed on: www.katinkavanderkooij.com/thesis/chapter2/abstractChapter2.html

In a discrimination task, observers could achieve good performance neglecting depth and formulating responses on the basis of other cues such as 2D relative velocity cues (Sperling et al., 1989). More pressing to the present purposes, such 2D velocity cues could interact between the centre and surround stimuli and observers would have different strategies at hand when incorporating information from the surround into their shape estimate. To keep observers from basing their responses on 2D velocity cues, rotation velocity was randomly varied by keeping the number of frames constant, but choosing the maximum rotation angle from an interval ranging from 12 º to 22º degrees back and forth. Whereas this manipulation was necessary to ensure that observers based their responses on 3D shape cues, it must be noted that perceived structure from motion is influenced by the rotation velocity (Domini & Caudek, 1999). The randomisation of rotation angle and velocity thus added noise to the comparison of the reference and test shape. To check whether this noise component qualitatively affected our results, we performed a control experiment where observers performed the conditions where the central shape was defined by motion without randomisation of rotation angle and velocity. Consistent with Domini & Caudek's results, shape estimates were more reliable compared to the shape estimates with randomised rotation velocity. But subjects displayed a wide range of surround-induced biases, which we attribute to the fact that they had different strategies at hand (using two-dimensional or three-dimensional information) when incorporating information from the surround in the shape estimate. Therefore, we continued the experiment with randomised rotation velocity.

Red and blue dots on the motion defined surfaces were projected into the cyclopean eye and these motion-defined shapes contained zero binocular disparity. In contrast, shape information from binocular disparity was created by projecting the stationary dots on the 3D hinged plane into the two eyes according to a perspective projection algorithm accounting for inter ocular distance. All stimuli were presented for 1.4 seconds.

Procedure

We presented ten different central test shapes for each of the four surround * four cue conditions. These 160 trial types were blocked by central shape cue (disparity or motion), and two successive blocks of the same cue were presented at a time, taking about 20 minutes. Each observer ran ten of such sessions, starting with a disparity block, which brings the total amount of measuring time per observer to about five hours, including breaks.

Observers first trained the angle discrimination task on stimuli where no surround was present with 13 replications of each test shape dihedral angle. Auditory feedback was given. After training, responses on the motion-defined shapes were compared to simulated responses based on rotation velocity. Participants that based their responses on rotation velocity, or with angle discrimination thresholds from motion or disparity larger than 30 degrees were excluded from further participation.

Participants

All observers had normal or corrected-to-normal visual acuity. They were seated in a dark room, 80 cm in front of an Iiyama Vision Master Pro 514 monitor with their heads in a chin rest to prevent head movements. Ten naive undergraduate students that passed the angle discrimination test participated, as did one of the authors (SP).

The psychometric curve

The point of subjective equality (PSE) and discrimination threshold were calculated by fitting the proportion of 'smaller dihedral angle' responses at each test stimulus angle. If there is no effect of the surround, the PSE should be equal to the angle of the central reference. Thus, the interesting parameter in this experiment is the difference between the PSE and the central reference angle: The bias, which we define in such a way that negative values represent contrast biases and positive value assimilation biases. We define the angle discrimination threshold as the 84% correct threshold that we obtain from the psychometric function.

Results

To test the hypothesis that slant contrast reflects integration of shape and surround by cue-dependent representations whereas assimilation reflects integration by cue-invariant representations, we compared biases in a 'within-cue' and 'cross-cue' condition. But first we checked whether there was an unexpected difference between the conditions where the dihedral angle of the test shape surround was 90 or 150 degrees. Bias data were entered into a repeated measures ANOVA with the factors Test Surround (90 or 150 degrees), Surround (same or different) and Cue Combination (within-cue or cross-cue). There was no effect of Test Surround angle, $F(1,10) = .21$, $p = .656$ for the condition where the central shape was defined by disparity and $F(1,10) = .21$, $p = 0.653$ for the condition where the central shape was defined by motion. Therefore we combined biases from these conditions.

Next, the data were entered into a repeated measures ANOVA with the factors Surround (same or different) and Cue Combination (within-cue or cross-cue). When the central shape was defined by disparity, there was an interaction of Surround (same or different) and Cue Combination (within-cue or cross-cue), $F(1,10) = 26.83$, $p < 0.000$ (Figure 2). We further looked into this interaction by a series of planned comparisons. When surround and shape were defined by the same cue, there was a *contrast* bias in the perception of the dihedral angle of the central shape; $t(10) = 4.79$, $p = 0.001$. But in the cross-cue condition where the inducers were defined by motion, *assimilation* of test and surround

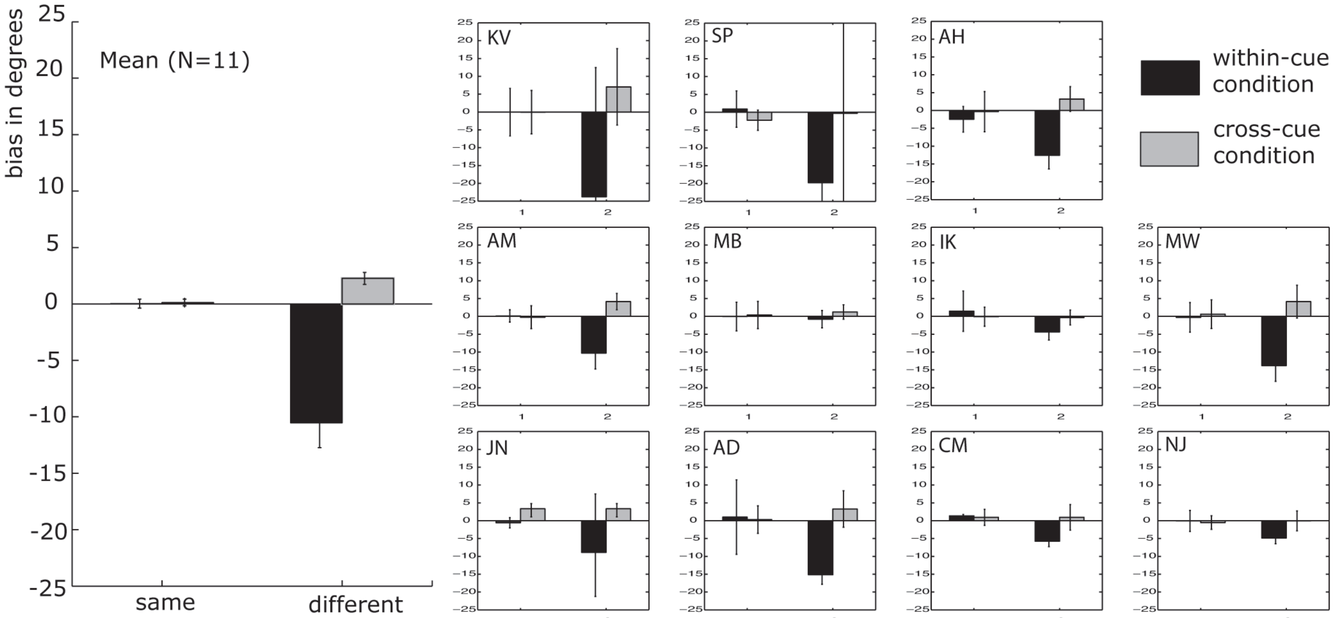
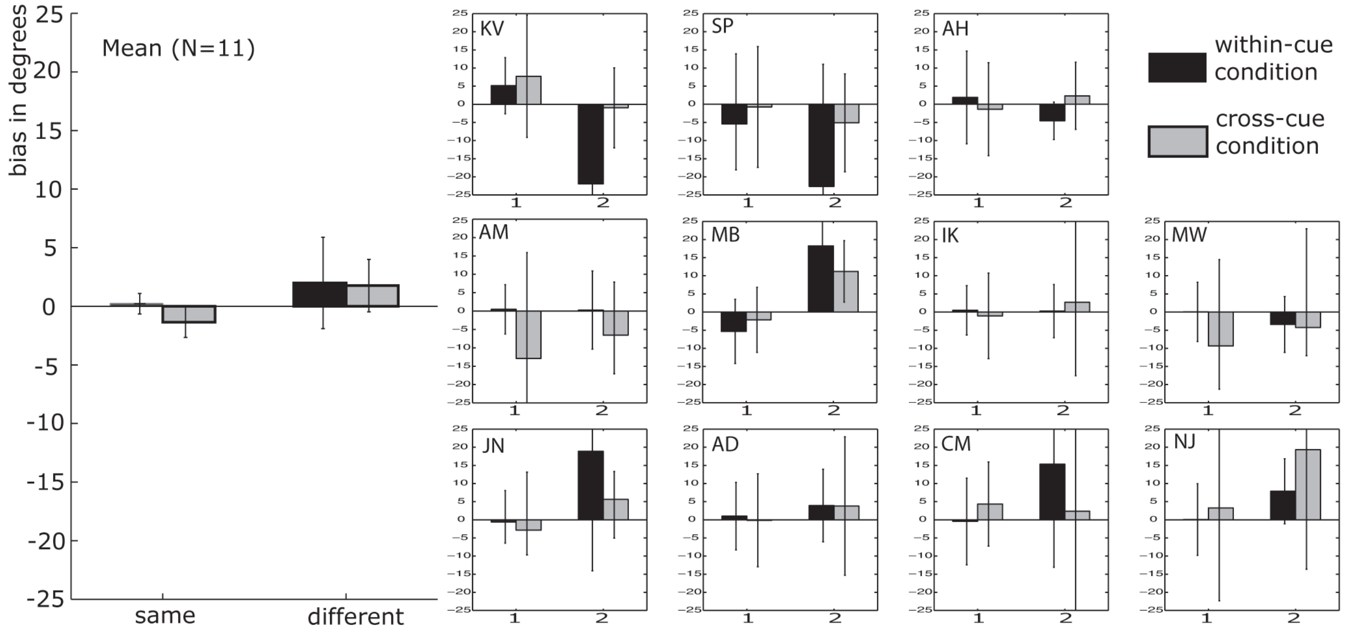


Figure 2. Bias in degrees of angle in the condition where the central shape was defined by disparity as a function of surround condition (same (1) or different (2)). On the left data averaged over 11 subjects. Error bars represent standard errors of the mean. Individual biases are presented on the right. Here, error bars represent the upper and lower limit of the confidence interval of the psychometric curve.



3 Figure 3. Bias in degrees of angle for the condition where the central shape was defined by motion as a function of surround condition (same (1) or different (2)). On the left data averaged over 11 subjects. Error bars represent standard errors of the mean. Individual biases are presented on the right. Here, error bars represent the upper and lower limit of the confidence interval of the psychometric curve.

dihedral angle occurred; $t(10) = -3.77$, $p = 0.004$.

When the central shape was defined by motion, there were no significant results. There was no effect of Surround ($F(1,10) = .59$, $p = 0.46$) nor interaction of Cue Combination and Surround ($F(1,10) =.19$, $p = .67$). But overall we observed two patterns of results (Figure 3). Three subjects (KV, AH and SP) showed a pattern of biases that was similar to the condition where the central shape was defined by disparity. When shape and surround were defined by the same depth cue, there was a contrast bias, whereas this was not the case in the cross-cue condition. The remaining eight subjects demonstrated an assimilation bias in both the within- and cross-cue conditions. This in contrast to the conditions where the central shape was defined by binocular disparity. Therefore, we looked for factors underlying these differences in bias direction.

Subj.	**DsameWC**	**DdifWC**	**DsameCC**	**DdifCC**	**MsameWC**	**MdifWC**	**MsameCC**	**MdifCC**
SP	9.57	14.14	11.19	7.85	34.46	48.50	38.41	45.56
AH	10.97	10.94	13.08	9.86	32.72	19.75	32.87	31.49
AM	6.30	9.34	6.99	5.72	25.72	35.64	47.24	42.34
CM	6.03	4.54	5.68	7.61	18.28	27.52	26.73	38.71
JN	6.26	8.21	7.65	6.83	25.15	41.13	36.64	26.75
KV	17.10	45.50	19.89	20.89	16.69	10.20	43.72	34.75
MB	4.25	7.90	5.31	5.50	35.18	31.46	39.22	26.65
NJ	7.64	5.80	6.10	8.78	21.04	26.29	50.22	49.91
IK	7.62	5.61	5.92	6.08	20.53	25.07	30.42	46.31
AD	13.34	5.78	11.62	11.06	28.56	30.69	35.95	48.75
MW	13.39	15.06	10.16	11.29	18.45	27.05	41.57	69.94

Table 1. Discrimination thresholds in degrees of angle for the individual subjects in the different cue and surround conditions (D = disparity, M = motion; same = same surround, dif = different surround; WC = within-cue

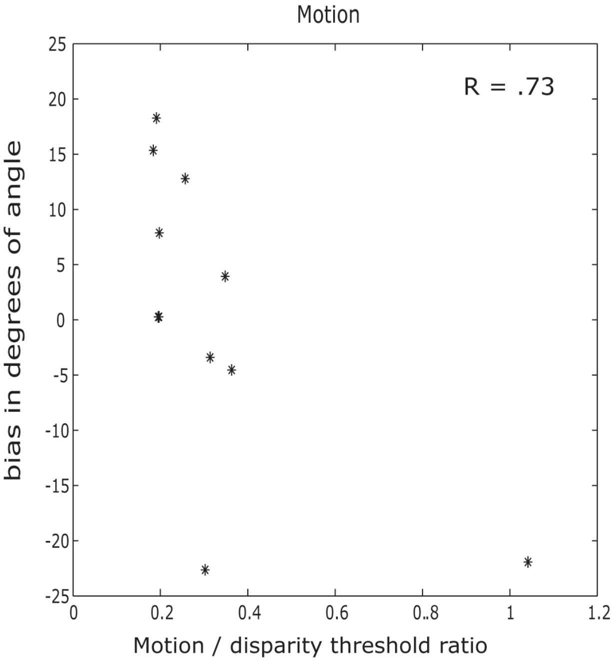


Figure 4. Correlation between the ratio of the motion and disparity threshold and bias (degrees of angle) in the motion within-cue condition.

condition, CC = cross-cue condition).

Subjects reported difficulty with the task and discrimination thresholds were much higher compared to the condition where the central shape was defined by disparity (with averages 10 degrees for disparity and 33.73 degrees for motion), *F*(3) = 48.08, *p* = 0.00. For some subjects the motion signal was unreliable to an extent where the psychometric curve did not fit properly, with unreliable estimates of the discrimination threshold as a result (Table 1). In itself, these unreliable estimates were unfit to predict biases. But signal reliability is estimated in the context of other shape signals and relative reliability might be even more telling. We compared the ratio of the discrimination thresholds for motion and disparity defined shapes, which can be taken as measure of the *relative* reliability of the motion cue, to the biases found in the different conditions. The ratio of the discrimination threshold for disparity and motion correlated well with the bias in the motion within-cue condition (*R* = .-743, *p* = .009). High motion/disparity threshold ratio's were associated contrast biases whereas low ratio's were associated with assimilation biases.

Discussion

We set out to answer the question whether both cue-dependent and cue-invariant representations are involved in the integration of shape and surround. To this end, we presented shape and surround in the same (the within-cue condition) or in a different depth cue (the cross-cue condition) and compared the surround bias between conditions. Shapes were defined by binocular disparity or motion.

When the central shape was defined by binocular disparity, *contrast* occurred in the within-cue condition whereas *assimilation* was observed in the cross-cue condition. When the central shape was defined by motion there were no significant results but two trends were observed. Some observers showed a pattern of results that was similar to the conditions where the central shape was defined by binocular disparity (i.e. contrast when all shapes were defined by motion, but assimilation when the surround was defined by disparity). But most demonstrated an assimilation bias not only in the cross-cue condition, but also in the within-cue condition where shape and surround were defined by motion. Crucial to our study, contrast occurred especially in the within-cue conditions, where integration could occur through cue-dependent representations. Therefore, we conclude that slant contrast reflects integration by cue-dependent representations whereas assimilation reflects integration by cue-invariant representations.

In contrast to our findings, van Ee et al. (1999) and Poom et al. (2007) explain both slant contrast and assimilation from a disparity-dependent mechanism, described in the Slant Estimation Model (van Ee et al., 1999). Both studies claim that slant is estimated from weighted linear combination of a direct shape estimate from cue combination and a disparity-based relative cue. The combination rule predicts slant contrast when shape and surround are defined by disparity but assimilation when the surround slant is defined by monocular cues whereas the disparity specified slant is zero. Van Ee et al. (1999) and Poom et al. (2007) only tested this cross-cue condition and their finding of assimilation in the cross-cue condition could be explained by the Slant Estimation Model. But the model predicts *enhanced* slant contrast when the central shape is defined by monocular cues and the surround by disparity. We show that in this case assimilation, not contrast occurs. With this, we are able to reject the Slant Estimation Model's position that assimilation is caused by weighted combination of a relative disparity-cue.

It is interesting to note that assimilation did not exclusively occur in the cross-cue conditions but also in the motion within-cue condition where all shapes were defined by motion. Bias size and direction (contrast or assimilation) were related to the ratio of the discrimination threshold for motion- and disparity-defined shapes, which can be taken as a measure of the relative reliability of the motion cue. Observers for whom the motion signal was relatively reliable (poor stereo viewers and good motion viewers), tended to show a contrast bias in the motion within-cue condition whereas observers to whom the motion cue was relatively unreliable tended to show assimilation biases in the motion within-cue condition. This finding is in line with the observation that centre-surround bias in motion perception shifts from contrast to assimilation with added visual noise (Hanada, 2004). In another study (van der Kooij and te Pas, 2009b) we directly assessed how a surround-induced bias depends on the reliability of shape signals by adding visual noise to the shape signals. This way, we showed that a contrast bias occurs when the shapes are well defined whereas an assimilation bias occurs when shape signals are unreliable. In short, the fact that we found an assimilation bias in the motion within-cue condition can be attributed to the difficulty subjects experienced with these shapes and we can maintain our conclusion that biases of shape contrast reflect integration through cue-dependent representations whereas assimilation biases reflect integration through cue-invariant representations.

But what mechanism might cause such assimilation? Recent neuro-imaging data have linked cue-dependent and invariant types of shape representation to areas of visual cortex and might offer a glimpse at the mechanism that causes assimilation of shape properties. Using fMRI adaptation methods, Welchman et al. (2005) showed that BOLD responses in early visual areas are related to the shape signal by individual depth cue, whereas the BOLD response in higher visual areas was related to the shape percept from cue combination. The information that can be represented in these areas is constrained by the receptive field properties of neurons within the area. Studies in cats and monkeys have shown that receptive field sizes are smallest in the central primary visual cortex (V1) and increase gradually in both higher and more peripheral parts of visual areas (Zeki, 1978, Van Essen et al., 1991, Maunsell and Newsome, 1987; Felleman and van Essen, 1991; Gattass et al., 2005). This means that cue-invariant representations are associated with larger receptive fields compared to cue-dependent representations. If the central shape and surround fell on a single population receptive field, the neural population would base its response on the average slant signal, which would explain the small assimilation bias in the conditions where integration had to occur through cue-invariant representations. In the case of unreliable information, the visual system can eliminate noise by averaging shape signals over a larger region, which might result in the use of more global representations (van der Kooij and te Pas, 2009b).

To conclude, both cue-dependent and cue-invariant representations are involved in the integration of shape and surround. A shape contrast bias reflects integration by cue-dependent representations whereas assimilation of shape properties is caused by integration through cue-invariant representations. Biases of shape contrast might be caused by relative shape cues, processed in early visual cortex whereas assimilation might be a by-product of large receptive field sizes of neurons in higher visual cortex where shape from cue-combination is processed.

Chapter 3

Uncertainty Reveals Surround Modulation of Shape

Published as: van der Kooij, K. & te Pas, S. F. (2009) Uncertainty reveals surround modulation of shape. *Journal of Vision, 3,* 9, p. 1-8

Abstract

Noisy estimations of shape can be partially resolved by incorporating relevant information from the context. The effect of surround stimuli on shape perception becomes clear in illusions of shape contrast and assimilation. In this study, we answer the question how a surround-induced bias depends on the reliability of shape signals. This way, we assess the processes by which an observer incorporates relevant data from the context into the shape estimate. We selectively added visual noise to the center and surround and compared a bias in shape perception with a control condition where no noise was added. In the conditions where shape and surround stimuli were well defined, we found a shape-contrast bias. When the surround stimuli were degraded, this contrast bias decreased. Most interestingly, when the central shape was degraded, an assimilation bias was observed. This bias was larger when the entire stimulus was degraded compared to when only the central shape was degraded. This suggests that shape contrast is the result of inference processes relying on local representations in early visual areas whereas assimilation is related to inference processes by global representations in higher visual areas.

Introduction

We rarely question the veridicality of our perception of the visual world but the inference processes involved occur under conditions of uncertainty. Three-dimensional (3D) shape, for example, has to be inferred from two-dimensional (2D) retinal projections where the mappings are not only one-to-many but also many-to-one. In such cases, noisy estimations can be partially resolved by incorporating relevant data from the context (Schwartz, Hsu, & Dayan, 2007). The influence of surround stimuli on perception becomes clear in biases of shape contrast, where perception is biased in the direction opposite to a neighboring stimulus. Such biases have been found in various geometric properties such as slant, curvature and depth and in stimuli that are defined by different depth cues, such as binocular disparity, motion and shading-and-texture (Anstis, 1975; Cornsweet, 1970; Curran & Johnston, 1996; Gibson, 1933; Graham & Rogers, 1982; te Pas & Kappers, 2001; te Pas, Rogers & Ledgeway, 2000). Assimilation, where shape perception is biased in the direction of the surround, also occurs but has been reported less frequently (van Ee et al, 1999; Poom et al., 2007; van der Kooij & Te Pas, submitted). Here we address the issue how surround stimuli influence the estimate of 3D shape.

Bayesian frameworks have been successfully applied to describe the perception of 3D shape from a combination of depth cues. In such an approach, an estimate of 3D shape is derived from weighted linear combination of shape estimates by depth cues such as binocular disparity, motion and perspective (Landy, Maloney, Johnston, & Johnston, 1995). Inherent in this kind of model is the idea that there are both cue-dependent (shape by cue x) and cue-invariant (shape from cue combination) representations of shape. This idea is supported by functional imaging data which show that BOLD responses in early visual areas are related to the shape signal from individual depth cues whereas activity in higher visual areas, such as LOC and MT+ is related to the combined shape estimate (Welchman et al., 2005). These representations might differ in more than cue invariance; receptive field sizes in the monkey visual cortex are much larger in these higher visual areas compared to striate cortex (Zeki, 1978). In a previous study (van der Kooij & te Pas, 2009a) we have shown that both types of representation are involved in the integration of shape and surround, causing qualitatively different surround-induced biases. When shape and surround were defined by the same depth cue, and integration could occur by cue-dependent representations, a contrast bias was observed. But when they did not share depth cue information, and integration had to occur through cue-invariant representations, an assimilation bias was found. In the latter case integration most likely had to occur through cue-invariant representations. Similar findings were reported by Poom et al. (2007) and van Ee, Banks and Backus (1999). Thus, an observer can integrate shape and surround by information from different types of representation and assign different weights to the information. The information carried by the representations might be constrained by the neural properties of different cortical areas.

Furthermore, the idea of weighted combination is supported by the findings that the percept depends on the reliability of information in the stimulus (Ernst & Banks, 2002).

But surround stimuli also provide different types of information to the observer. They provide relative shape cues (i.e. the difference between surround and central shape) that have to be weighted (van Ee, Banks and Backus, 1999). If a high weight is given to this difference, perception might become biased in the direction opposite to neighboring stimuli. In line with such reasoning, a reduced contrast bias in slant perception has been found with an increased gap between shape and surround, which reduces the reliability and consequently the weight of relative shape cues (Poom, Olsson, & Borjesson, 2007). For these relative shape cues to be effective, shape and surround have to fall on different receptive fields. On the other hand, there are substantial correlations between image points that are spatially near to each other or that are close in time. This means that the temporal and spatial context will induce expectations about the value in a certain image point (Howe & Purves, 2005). Observer's knowledge of such correlations is evident in their ability to replace missing pixels in digital images based only on neighborhood information (Kersten, 1987).

In conclusion, an observer can rely on different types of shape representation and has different types of information at hand when incorporating information from surround stimuli into the shape estimate, which may lead to biases of shape contrast or assimilation. Recent findings (Poom et al., 2007; van der Kooij & te Pas, submitted) suggest that a contrast bias is the result of integration by cue-dependent representations whereas an assimilation bias is related to integration by cue-invariant representations, but the issue how surround stimuli influence the perception of 3D shape can only be further resolved after it has been clarified what causes the shift in bias direction (contrast or assimilation).

Because the outcome of Bayesian inference depends strongly on the reliability of shape signals, external noise methods have successfully been used as a tool for system identification (Ernst & Banks, 2002; Tjan, Lestou, & Kourtzi, 2006). From this given, we develop a simple paradigm to show how surround-induced bias direction depends on the reliability of shape signals. Thereby we illuminate the inference processes by which information from surround stimuli is incorporated in the 3D shape estimate. Relative shape cues become less reliable and loose influence on perception when surround information is degraded. Therefore, a contrast bias will be decreased. But when the shape signal of the central stimulus is degraded, local information might be impoverished to the extent where the observer comes to rely on more global shape representations. Integration by these representations might cause averaging of the shape signal from the central shape and surround and an assimilation bias in shape perception. Also, an observer might come to rely on a different type of information when integrating information from the surround in the shape estimate.

In summary, we expect illusions of shape contrast to decrease when shape stimuli are degraded whereas we expect assimilation biases to become visible when the shape stimuli are degraded. Although the dependency of shape-surround interactions on the reliability of shape signals is relatively unexplored, studies on center-surround interactions in motion provide results that are in line with our predictions for shape (Hanada, 2004).

We reduce the reliability of shape signals by adding local noise to the shape stimuli (Ernst & Banks, 2002; Tjan et al., 2006). Noise is selectively added to the center and surround stimuli and surround-induced shape biases are compared to a control condition where no noise is added. In this control condition, we expect to replicate the findings of a slant contrast effect. In the condition where the surround is degraded, we expect relative shape cues to receive less weight and a decrease in bias. But of greatest interest are the conditions where the central shape is degraded by added noise. In this case we expect an assimilation bias in shape perception.

Methods

Random dot stereograms depicted a hinged plane receding in depth, flanked by two larger hinged planes (Figure 1). Random dots were back-projected from the screen onto the 3D structures in such a way that texture and perspective cues to shape would signal a flat surface. All shapes were 21.6° visual angle high, the central shape was 5.9° wide and flankers measured 12.9° of visual angle in width. Central shape and surround stimuli were separated by a gap of 1° visual angle. Pixel density was 2.7 pixels per degree of visual angle and one pixel subtended 0.001 degrees of visual angle. Observers made dihedral angle comparisons of a range (60° to 140°) with a constant reference of 100° dihedral angle (method of constant stimuli).

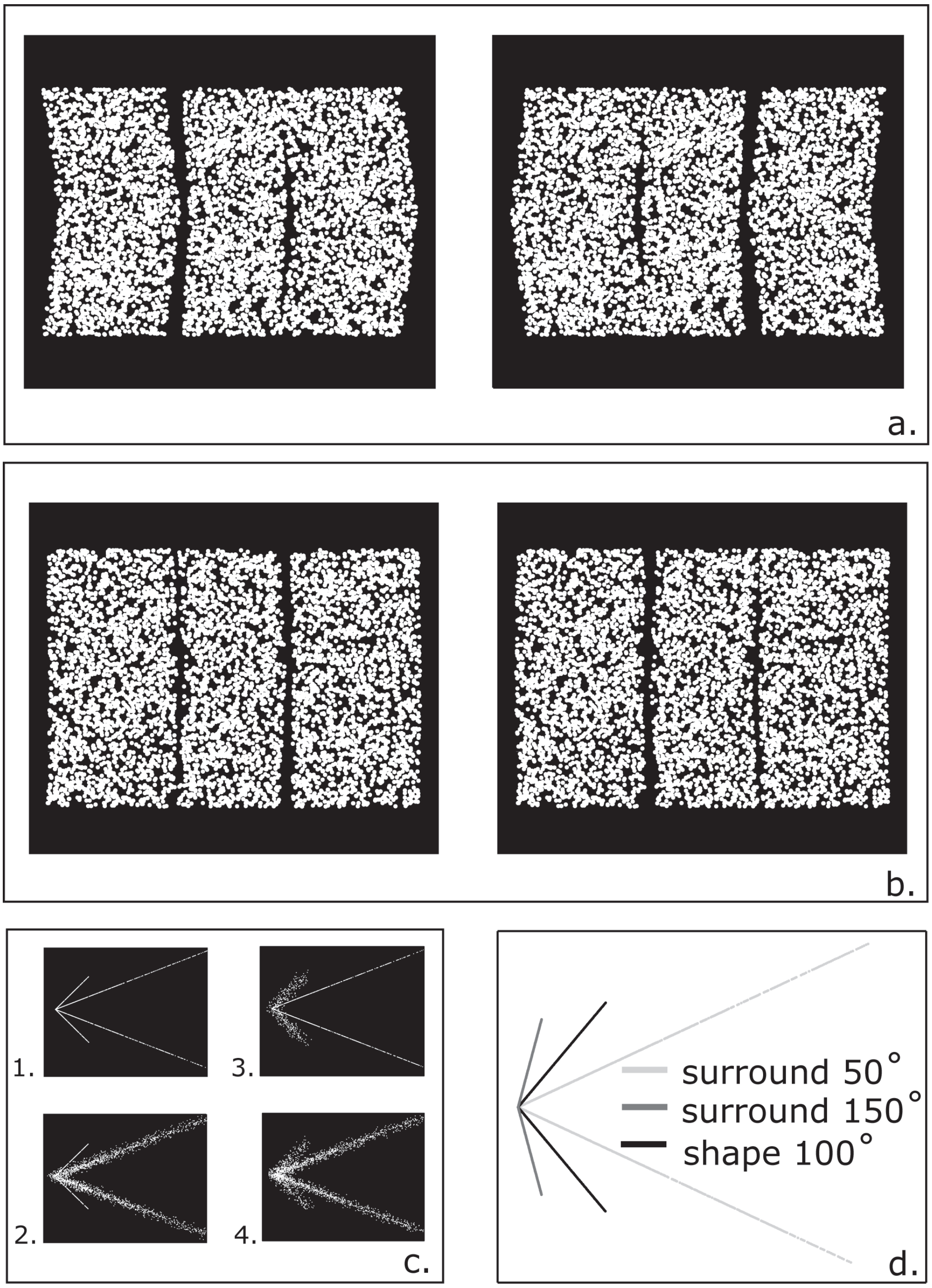


Figure 1. Stimulus set-up. a. stereogram of the condition where the dihedral angle of the surround was 50°. b. stereogram of the condition where the dihedral angle of the surround was 150°. Both prepared for fusion. c. noise conditions, (1) 'no noise', (2) 'surround noise', (3) 'target noise' and (4) 'stimulus noise' condition for the situation where the surround had a dihedral angle of 150°. c. Side-view cartoon of the shape stimuli (the separation of the x-axis is not visible in this side-view). Anaglyph demo's can be viewed on: www.katinkavanderkooij.com/thesis/chapter3/abstractChapter3.html

To measure a surround bias, we varied the dihedral angle of the surround between the reference and test interval. The dihedral angle of the surround stimuli was either 50° or 150° and there were two surround conditions: one where the dihedral angle of test surround was smaller compared to the reference surround and one where it was larger. Shape reliability was varied by selectively adding local shape noise to the stimuli. Local shape noise was created by displacing dots on the shape surface in a random direction in three-dimensional space, resulting in a maximal displacement of 3.6 centimeters orthogonal to the hinged plane. In a 'surround noise' condition, noise was selectively added to the surround shape stimuli, in a 'target noise' condition noise was added to the central target and in a 'stimulus noise' condition the entire stimulus was degraded and in a control condition no noise was added (Figure 1.c).

Procedure

The two surround conditions ('larger' or 'smaller') x 4 noise conditions ('no noise', 'surround noise', 'centre noise' or 'stimulus noise') x 16 test dihedral angles resulted in 128 different trial types, of which subjects performed 10 replications. A trial started with a 750 ms presentation of a central shape and surround, followed by a 750 ms gray fixation cross and the presentation of a second shape and surround for 750 ms. Next, the fixation cross reappeared and the observer decided whether the first or second central shape had the smallest dihedral angle. After the observer responded the fixation-cross turned blue to indicate that a key-press had been given and stayed on the screen for an inter-trial interval of 1 second. As noise is estimated over a period of several trials (Hanada, 2004), these trials were blocked by noise condition, and each block took about 30 minutes to measure. All subjects started with the 'no noise' condition, followed by the 'surround noise' condition and continued with the 'stimulus noise' and 'target noise' conditions. This way, lower discrimination thresholds in the conditions where no noise was added to the central shape could not be attributed to a learning effect. To assess subjects' stereo vision and to further minimize effects of learning during the experimental conditions, subjects first trained angle discrimination from binocular disparity with 10 replications of each test dihedral angle on stimuli where only the central shape was presented. Feedback was given by a change of fixation-cross color (green for correct and red for incorrect). In total, the training phase took 20 minutes. Five subjects who's stereo vision was too poor to perform the task and for whom the psychometric curve could not be fitted in either of the conditions were excluded from further participation. Two of the authors and four subjects that were naive as to the purposes of the experiment participated. All subjects had normal or corrected-to-normal vision.

Analyses

The PSE and discrimination threshold were calculated by fitting the proportion of 'smaller dihedral angle' responses at each test stimulus angle with a cumulative Gaussian (the psychometric function, see Figure 2). The angle discrimination threshold, which we define as the 84% correct threshold that we obtain from the psychometric function, results in a measure of shape reliability (1/discrimination threshold). In the absence of the surround, the PSE should be equal to the angle of the central reference. Thus, the interesting parameter in this experiment is the difference between the PSE and the central reference angle: the bias. We define the bias in such a way that negative values represent a shift away from the surround dihedral angle (contrast) and positive values represent a shift towards the surround dihedral angle (assimilation).

3.3 Results

To test how a surround-induced bias in shape perception depends on visual noise, we compared the biases and discrimination thresholds from the individual subjects (Table 1) between noise conditions with a repeated measures ANOVA (Figure 3). We first checked for differences between the conditions where the dihedral angle of the test surround was smaller or larger compared to the reference

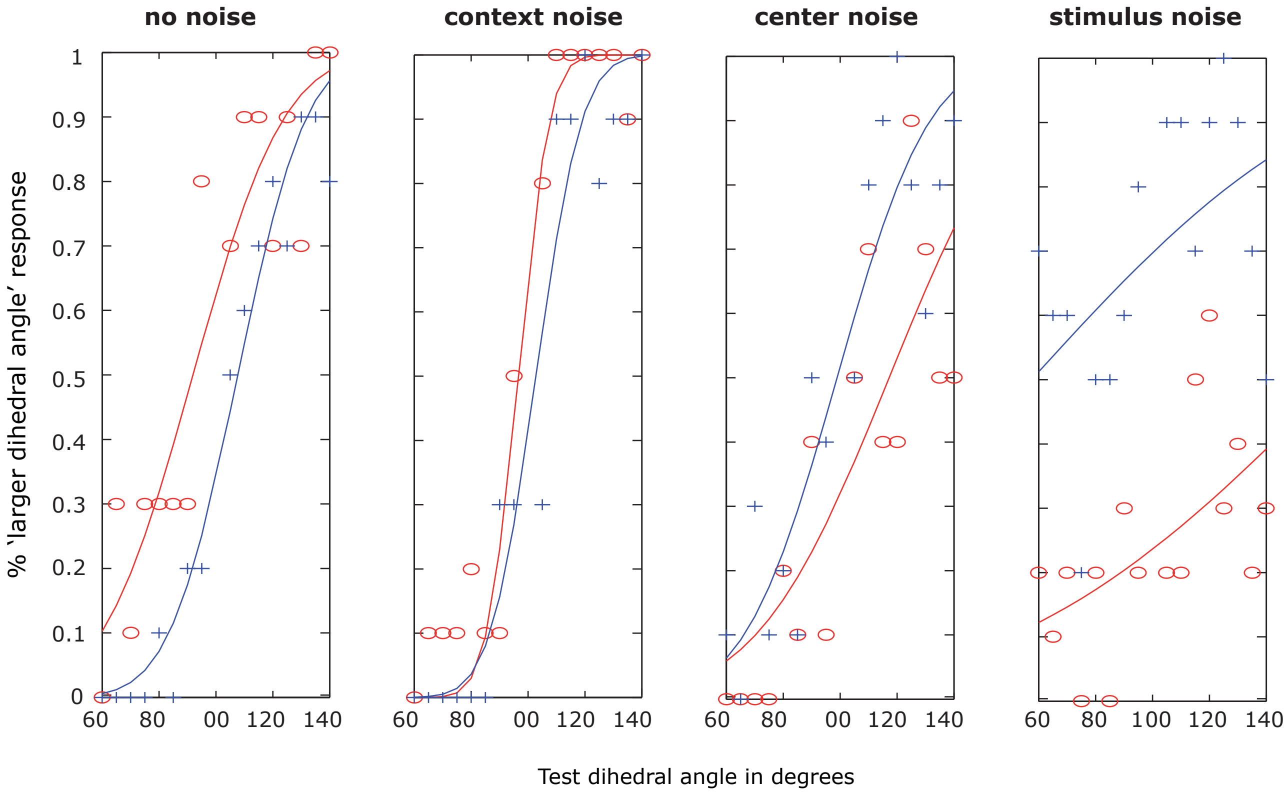


Figure 2. The fitted psychometric curves for the different surround and noise conditions on the data from an example subject (HN).

surround dihedral angle. A repeated measures ANOVA with the factors Surround ('larger dihedral angle' or 'smaller dihedral angle') and Noise ('no noise', 'surround noise', 'target noise' or 'stimulus noise') on the bias and discrimination threshold data from the 6 subjects revealed no differences in bias ($F(1,5) = 0.06$, $p = 0.82$) or discrimination threshold ($F(1,5) = 0.12$, $p = 0.74$) and we combined data from the two conditions. Veridicality of shape perception without surroundings was assessed by an analysis of the training data where no surround stimuli were present. In both cases average biases did not differ significantly from zero.

Next, we compared bias size between noise conditions. There was a significant effect of Noise condition ($F(5) = 4.36$, $p = 0.02$). We further looked into this effect with a series of one-tailed paired samples t-tests. Whereas we found an contrast bias in the 'no noise' condition, an assimilation bias was observed in the 'stimulus noise' condition ($t(5) = -3.74$, $p = 0.03$). Furthermore, bias size was smaller in the 'surround noise' condition compared to the 'no noise' condition ($t(5) = -2.44$, $p < 0.01$). There seemed to be a similar decrease in bias size between the Target Noise and Stimulus Noise conditions, although this did not reach significance. One subject, SP, deviated from this pattern and showed a contrast bias in both the 'no noise and stimulus' noise conditions. We attribute this to the fact that she was a highly experienced psychophysical observer and might have developed different strategies in estimating shape.

As the discrimination threshold is inversely related to shape reliability, we next compared discrimination thresholds between noise conditions with a repeated measures ANOVA with the factor Noise ('no noise', 'surround noise', 'target noise', 'stimulus noise'). There was a main effect of the factor Noise ($F(5) = 9.57$, $p = 0.001$). Further investigation of this effect with a series of one-tailed paired-samples t-tests showed that whereas discrimination thresholds were much higher in the conditions where noise was added to the central target compared to the conditions where no noise was added to the target, there was no significant difference between the noise conditions within these two situations ('no noise' vs. 'surround noise' ($t(5) = 1.71$, $p = 0.43$) and 'target noise vs. 'stimulus noise' ($t(5) = -1.73$, $p = 0.07$) (Fig. 3)

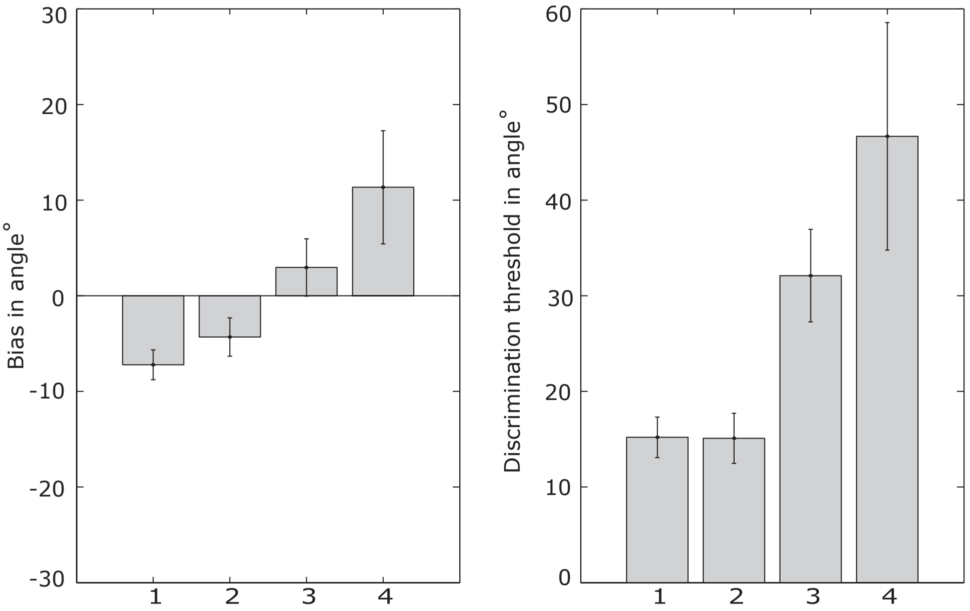


Figure 3. Average bias and discrimination threshold, derived from the psychometric curve, in degrees of angle as a function of noise condition (1 = 'no noise', 2 = 'surround noise', 3 = 'target noise', 4 = 'stimulus noise'. Negative biases mean contrast and positive biases mean assimilation. Error bars represent standard errors of the mean (N = 6).

Subject	**No noise**		**Surround noise**		**Target noise**		**Stimulus noise**	
	Bias	**T**	**Bias**	**T**	**Bias**	**T**	**Bias**	**T**
SM	-10.80	19.56	-11.57	22.75	12.12	68.13	13.30	70.30
HN	-7.89	22.03	-3.19	10.71	9.16	31.09	53.54	86.06
SP	-4.37	17.30	-9.13	19.74	-27.02	41.18	-19.90	38.53
TG	-4.37	8.90	-0.78	7.26	17.49	15.0	21.72	20.14
KK	-3.04	12.29	-0.42	18.62	15.63	25.87	30.23	37.38
JV	-6.67	12.21	-2.53	10.66	1.08	19.53	31.08	37.16

Table 1. Bias and discrimination threshold (T) for each of the subjects in degrees of angle observed in the different noise conditions.

Discussion

Unreliable estimates of shape can be partially resolved by incorporating relevant information from the surround in the shape estimate. This could occur through different types of shape representation. Also, an observer could use different strategies. In this paper, we address the issue how information from surround stimuli is integrated in the estimate of 3D shape. As, external noise methods have proved a powerful tool for system identification (Ernst & Banks, 2002) we answer the question how a surround-induced perceptual bias depends on the reliability of shape signals. To this end, we selectively added correlated shape noise to the central shape or surround and compared a surround bias with a control condition where no noise was added to the stimuli. Overall, a reliable central shape was associated with a contrast bias whereas an unreliable central shape was associated with assimilation biases. In the control condition, we replicated findings of a contrast bias in shape perception (Poom et al., 2007;

te Pas & Kappers, 2001; van Ee et al., 1999). When only the surround was degraded, the contrast bias was diminished and when only the central shape was degraded, the assimilation bias tended to diminish, although this did not reach significance. This suggests that the reliability of the central shape determined bias direction, whereas the reliability of the difference between the central shape and surround determined bias size. Yet, it is important to note that low reliability of the central shape is not a necessary condition for assimilation. Several reports exist of assimilation biases between hinged planes when the central shape and surround were defined by a different depth cue (van Ee et al., 1999; van der Kooij & te Pas, submitted; Poom et al., 2007). In all studies discrimination thresholds for the central shape were much below the discrimination thresholds for the noisy surfaces in this study.

The way in which a surround induced bias depends on the reliability of shape signals could clarify whether biases of shape contrast and assimilation are caused by a single mechanism or whether they are the result of complementary processes. Whereas a contrast bias can be explained with relative shape cues, which are combined with direct shape cues (van Ee et al., 1999), Bayesian models of shape contrast cannot explain the shift in bias direction with added noise. We will first show how two highly intuitive hypotheses, which explain contrast and assimilation effects from a single mechanism, cannot account for the full set of our results. Next, we show how the evidence presented here relates to the neural architecture involved in the perception of three-dimensional shape and make a case for the position that biases of shape contrast and assimilation are the result of complementary processes. First, one might attribute the shift in bias direction to the fact that adding shape noise to the central shape or surround diminished the reliability of the difference between the central shape and surround up to an extent where the difference between shape and surround is below threshold. Contrast would be the result of the perception of two Gestalten whereas assimilation results from the perception of one Gestalt (van Lier & Wagemans, 1997). But this explanation cannot account for the full set of our results. The area between the central shape and surround surface can be taken as a measure of the reliability of the difference between shape and surround. Using this measure, the reliability of the difference between shape and surround is about equal when only the central shape is degraded or when only the surround is degraded. But perceptual biases are very different between the two conditions: when the surround is degraded, contrast occurs, but when the central shape is degraded, assimilation is observed. Alternatively, one could propose that the surfaces to which noise was added appeared to have less depth and, therefore, larger dihedral angles. When the surround has a smaller dihedral angle compared to the central shape, this would have indeed decreased the difference between the central shape and surround. But in the situation where the surround has a larger dihedral angle, the difference between central shape and surround would be enhanced. As we tested both situations, these shifts in bias would have averaged out. To summarize, the shift in bias direction cannot be explained from low reliability of the difference between shape and surround, nor can it be explained by 'flattening' of dihedral angles with added noise.

Instead, we make a case for the position that contrast and assimilation biases are the result of complementary processes that rely on shape representations in different cortical areas. Both cue-dependent and cue-invariant representations are involved in the perception of 3D shape and surround stimuli can influence the shape estimate through both types of representation (van der Kooij & te Pas, submitted). Cue-dependent representations have been linked to activity in early, striate, visual areas whereas cue-invariant representations have been linked to fMRI activity in higher visual areas such as the lateral and temporal occipital cortex (Welchman et al., 2005). Furthermore, Studies in cats and monkeys have shown that receptive field sizes are smallest in the central primary visual cortex (V1) and increase gradually in both higher and more peripheral parts of visual areas (Zeki, 1978, Van Essen et al., 1984, Maunsell and Newsome, 1987; Felleman and van Essen, 1991). Estimates range from less than 1 degree in V1 to more than 8.5 degrees in LOC. This means that our central shape is much larger than receptive field sizes in early visual areas where information from individual depth cues is processed whereas the entire stimulus, encompassing the central shape and its surround, falls into the receptive fields of higher visual areas. We attribute biases of shape contrast to relative shape cues. For such relative cues to be effective the central shape and surround must fall on different receptive fields, so that local information can be compared. For this task, the relatively small receptive fields in early visual areas would especially be effective. Assimilation, on the other hand, might have been caused by shape and surround falling on a single receptive field in higher visual areas such as LOC and the temporal occipital cortex. If neurons in this area are unable to differentiate between shape and surround they would average the depth signal over the entire receptive field, causing an assimilation bias. Besides being physiologically plausible, this explanation of contrast versus assimilation biases is also in accordance with reports that assimilation biases are associated with integration by cue-invariant representations (in higher visual areas) whereas contrast biases are associated with integration by cue-dependent representations, typically found in early visual areas. The switch from contrast to assimilation with added noise might be caused by the fact that our noise methods especially degraded local information, shape could only be perceived after averaging depth information over the shape surface. Therefore the observers might have come to disrespect local information in favor of global information, probably represented in higher visual areas with large receptive fields. Additionally, when shape information is unreliable, the observer might also take into account the correlations that exist between image points that are spatially near to each other. Observers' knowledge of such correlations is evident in their ability to replace missing pixels in digital images based only on neighborhood information (Kersten, 1987). If observers fill in unreliable shape information with information from the surround by averaging or spatial smoothing, biases of assimilation would easily occur.

To summarize, we hypothesize that biases of shape contrast are caused by weighting of relative local shape cues, represented in early visual areas where receptive visual fields are small, whereas assimilation biases are caused by the low spatial resolution of receptive fields in higher visual areas where 3D shape from cue combination is encoded. When perceptual information in the central target is poor, subjects might come to rely more on the global shape signal in these higher areas.

3.5 Conclusions

We have shown that bias direction in shape-surround interactions depends on the reliability of shape signals. When shape information is relatively reliable, the surround invokes a contrast bias in shape perception. But when noise is added to the entire stimulus assimilation of shape and surround occurs. These findings suggest that shape and surround are integrated at different levels of the neural architecture involved in the perception of 3D shape. Relative information might be represented in lower visual areas whereas global shape information might be represented in higher visual areas such as the lateral and temporal occipital cortex.

Chapter 4

Shape Contrast: A Global Mechanism?

In Preparation at Vision Research as:
Van der Kooij & S. F. Te Pas (2010) Shape Contrast: A Global Mechanism?

Abstract

We investigated whether a shape contrast bias is caused by local contrast enhancement or by a global mechanism. In a baseline condition, observers performed a shape discrimination task on an isolated hinged plane. But in the experimental conditions, five dihedral surfaces, of which we varied the dihedral angle distribution, were added on each side. Shape perception was influenced not only by the adjacent surface but also by the mean of the shape distribution in the extended surround. Thus, shape contrast is not locally determined and has to be understood from a global mechanism. We propose divisive normalization of shape signals as such a mechanism.

Introduction

Most psychophysical studies have reduced the heterogeneity of visual stimuli for reasons of simplicity. But even when stimulated with only two discrete stimuli, perception has been found to depend heavily on spatial context (Schwartz, Hsu, & Dayan, 2007). For example, when a fronto-parallel plane is viewed between two backward slanted planes, it is perceived to be slanted forward (e.g. Graham & Rogers, 1982). In this study, we address the issue of how the visual system processes shape information over a large area of the visual field. To this end, we answer the question whether a shape contrast bias is affected not only by interactions between neighboring shapes but also by the shape distribution over an extended area of the visual field. Shape contrast biases are very robust, yet the mechanisms that produce contextual biases in shape perception remain obscure. Two types of mechanism have been proposed, which we will briefly introduce below, before discussing how these mechanisms relate to the spatial extent of shape contrast.

The first type of mechanism has been proposed based on the fact that contextual stimuli provide relative information (‘cues’) about shape. This relative information might be integrated with absolute information about shape to increase the reliability of shape judgments (van Ee, Banks, & Backus, 1999). The problem of how the visual system integrates different types of information about shape can be described in a maximum likelihood estimation (MLE) mechanism (e.g. Knill & Saunders, 2003; Landy, Maloney, Johnston, & Johnston, 1995). In such a mechanism, shape information is integrated according to its reliability to determine the most likely underlying source. This means that unreliable information will have less influence on the final estimate. In an MLE framework of shape contrast, contrast biases are thought to arise because the visual system is more sensitive to relative than to absolute information (Gillam, Blackburn, & Brooks, 2007; van Ee, Banks & Backus, 1999). Because relative information is the more reliable cue, it will have a large influence on the shape estimate. However, the exact way in which relative and absolute information are combined, remains unknown. Thus, in an MLE mechanism of shape contrast, contrast biases arise from improper combination of relative and absolute information about shape.

The second type of mechanism has been proposed based on the observation that contextual effects in perception have a neural counterpart. In ‘center-surround’ effects, neurons coding for similar features can suppress or facilitate each others response to visual stimulation (e.g. Cavanaugh, Bair, & Movshon, 2002). Suppression of responses to similar stimuli causes a contrast bias in the population tuning curve from which slant is decoded, which might explain the contrast bias in slant perception (Schwartz, Sejnowski, & Dayan, 2009). Neural models have shown that surround suppression can be described with a divisive normalization operation where the common factor in a stimulus is averaged out (e.g. Cavanaugh et al., 2002; Heeger, 1992; Kouh & Poggio, 2010). Therefore, we refer to this type of mechanism, operating on surround suppression as a divisive normalization mechanism. On the functional level, divisive normalization has the advantage that redundancy in visual information is reduced (Schwartz & Simoncelli, 2001).

Both MLE and divisive normalization mechanisms have been proposed to explain interactions between adjacent shapes, but may also involve shape interactions on a global level, as we will explain below. MLE models have been very successful in describing how information within a 3D surface is combined to estimate its shape (e.g. Knill & Saunders, 2003; Landy et al., 1995), but little is known about MLE integration of signals coming from different surfaces. Essentially, MLE integration of visual information only makes sense when information has a common source but indirect global effects may be possible. One specific MLE model of shape contrast, proposes that shape contrast biases arise because the contextual surface serves as a reference against which relative differences are perceived (van Ee et al., 1999). This means that before combining relative and absolute information, the relative information has first been gauged to a reference. In the specific model by van Ee and others (1999), the reference consists of the flankers, but in theory, relative differences could also be gauged to a larger reference plane, or the average shape in the stimulus.

Just as MLE models, divisive normalization of visual signals has been well established on the local level, whereas little is known about divisive normalization in complex stimuli. On the neural level, divisive normalization may be accomplished by different mechanisms. A likely candidate are the horizontal connections between neurons coding for similar features and which can link regions over several millimeters (Series, Lorenceau, & Fregnac, 2003). But normalization may also be accomplished by feedback from higher areas. First, the speed with which information travels over horizontal connections is not sufficient to account for the time course of normalization (Bair, Cavanaugh, & Movshon, 2003). Second, blocking the transmission of activity through horizontal connections does not fully disrupt normalization (Brown, Allision, Samonds, & Bonds, 2003). Because higher areas typically code for more global aspects of a scene, normalization through feedback connections could take into account global properties of a scene. In addition, local interactions through horizontal connections may add up to a global effect. Furthermore, divisive normalization may conceptually be consistent with a more global mechanism. If the visual system normalizes visual responses to reduce redundancy, it makes most sense to average out the common factor over a relatively large region.

To conclude, both MLE and normalization mechanisms have been well established on the local level, but do not rule out shape interactions on a global level. The main difference between a global MLE mechanism and global divisive normalization, is that in an MLE mechanism, the contrast bias will depend both on the variance and mean of the shape distribution in a scene whereas in a divisive normalization mechanism, the shape contrast bias will depend only on the mean of the shape distribution.

To test how a shape contrast bias depends on the shape distribution over an extended area, observers performed a shape discrimination task on two sequentially presented hinged planes, which were defined by disparity and flanked by a five other hinged planes on each side. We kept the dihedral angle difference constant between the central shape and neighboring surfaces, but varied the angle distribution of the extended surround.

Methods

Observers

Eight observers, including one of the authors, a graduate student and six bachelor students at Utrecht University participated. Students received course credit or a financial reward of 6 euro per hour. All observers had corrected, or corrected to normal vision.

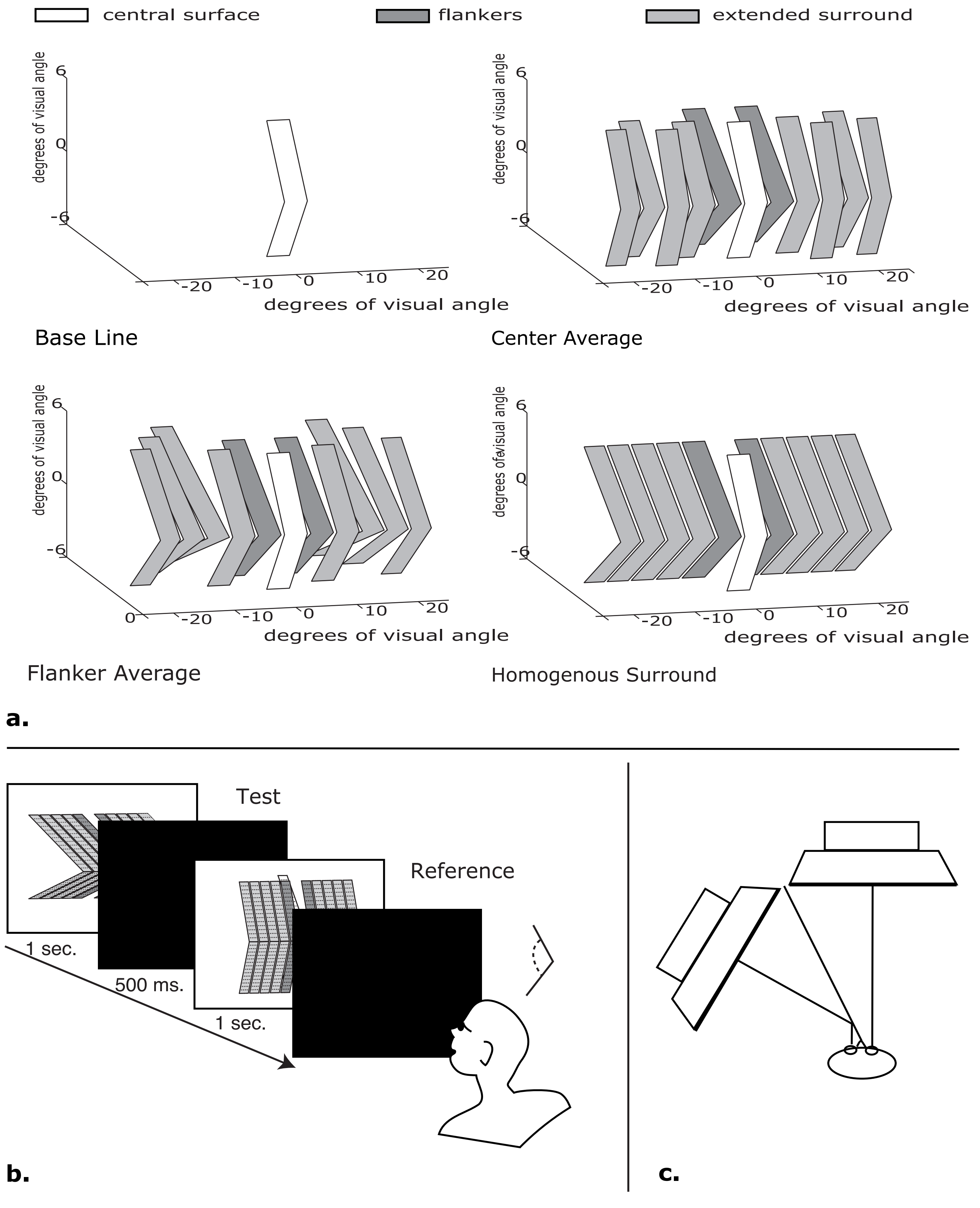


Figure 1. a) Stimulus lay-out in the different surround conditions. **b)** Trial sequence in the angle discrimination task. **c)** Cartoon of the stereo set up, where the right eye viewed the right monitor directly and the left eye viewed the left monitor through a mirror placed on the diagonal between the two screens. Anaglyph demonstrations of the stimuli can be viewed on: www.katinkavanderkooij.com/thesis/chapter4/abstractChapter4.html

Stimuli & task

Stereograms were viewed on a mirror set-up where the right eye viewed the right monitor directly and the left eye viewed the left monitor through a mirror on the diagonal between the two monitors (Figure 1.2). Stimuli were presented on two Brilliance 202P7 Philips CRT monitors (1920 x 1440, 75 Hz). Square random dots measured 1.5 arcmin visual angle and had maximal luminance contrast.

Random dot stereograms depicted a hinged plane receding in depth and five more hinged planes on each side, with the creae of the surfaces aligned at screen depth (Figure 1.1). The screen projection of all shapes was 2.5º visual angle wide and 7.5º visual angle high. As surfaces were horizontally separated by a 0.4º visual angle strip of black screen, this resulted in the total stimulus measuring 31º in width. We call the neighboring surfaces the 'flankers' and the eight exterior most surfaces the 'extended surround.'

Depth information in the screen image was created by perspective projections. To be able to place the 3D surfaces in close vicinity without having their projections overlap, rectangular random dot patches were back-projected from the screen onto a hypothetical 3D structure of hinged planes. Next, binocular disparity was created by projecting the random dots from the 3D structure into the left and right eye. This way, texture and perspective cues to depth were minimized.
As observers were free to make eye movements over the stimulus, we used parallel lines of sight where the eyes convergence on a hypothetical viewpoint at infinite distance away from the surfaces. To further ensure that for no subject the projections of the different surfaces overlapped, we chose an eye-distance of 6 cm. for all observers. An eye-distance of 6 cm would have resulted in a 7 degree convergence angle on screen depth.

Pixel density within a surface was 2 %. To measure shape perception of the central surface, observers performed a sequential dihedral angle comparison task where they compared the dihedral angle of a test surface, ranging from 64 to 136 degrees, with a constant reference of 100 degrees. After an inter-trial-interval of one second, a random order of the test and reference surface, separated by an inter-stimulus-interval of 500 ms, was presented. Test and reference shape were each presented for one second and to maintain fusion and fixation on the center of the screen, a fixation cross was presented during the inter-trial and inter-stimulus intervals. After the two shape stimuli had disappeared, the observer indicated, using a key-press, whether the central surface in the first or second interval contained the largest dihedral angle (Figure 1.b).

To induce a shape contrast bias, we created two flanker conditions: one where the flankers had a dihedral angle of 60 degrees and another where the flankers had a dihedral angle of 140 degrees. These angle differences induced maximal contrast biases in a pilot study. To investigate the effect of the extended surround on shape perception, we created three surround conditions by varying its shape distribution. In a 'homogenous surround' condition, the surfaces in the extended surround had the same dihedral angle as the flankers. But in the 'center average' and 'flanker average' conditions, variance was added to the shape distribution by randomly choosing each surface in the extended surround from a uniform distribution with a range of 40 degrees. The left and right section of the stimulus were mirror symmetric. In the center average condition, the mean of the distribution from which surround angles were chosen was equal to the central angle, whereas in the flanker average condition the mean was equal to the flanker angle (Figure 1.a). Angle discrimination in these surround conditions was compared to a baseline condition where only the central surface was presented.

Staircase procedure

We retrieved the point of subjective equality (PSE) for dihedral angle perception using a one-up, one-down staircase procedure. Staircases started with a comparison stimulus with a dihedral angle that was 36 degrees smaller or larger compared to the reference stimulus. A staircase starting with a negative difference and a staircase starting with a positive difference were interleaved. These staircases terminated after ten reversals. Additionally, staircases were terminated when the test angle was smaller than ten degrees. For these sharp folding angles, observers lost stereo-vision due to the large cue conflict between disparity signaling a sharp folding angle and other cues such as perspective, texture and accommodation cues, signaling a flat surface. To obtain a measure of the PSE and discrimination threshold, responses at each test stimulus angle, derived from the staircase procedure, were fitted with a cumulative Gaussian. The 84% correct threshold is the angle discrimination threshold, and the PSE is the 50% correct point. The difference between the PSE and the reference angle is taken as the bias and is signed in such a way that positive values represent a shift away from the flanker angle (contrast) and negative values represent a shift towards the flanker angle.

Procedure

Trials were blocked for surround condition and all observers started with the baseline condition, where only the central shape was presented. If standard deviation of the PSE measured in this condition exceeded fifteen degrees of dihedral angle, we took this as a sign of poor stereovision and excluded the observer from further participation. This resulted in the exclusion of three subjects. Next, the observers performed a random order of the three surround conditions, which each took about ten minutes of measuring time. To further minimize the effect of test order, observers came back on a different day for a second session in which they performed a different order of the surround conditions, again starting with the baseline condition. Data were averaged over the two sessions.

Results

We obtained the bias by subtracting the PSE in the baseline condition from the PSE's in the different Flanker and Surround conditions and scoring the data such that positive values represent as bias away from the flankers (contrast) whereas negative values represent a bias towards the flankers (assimilation).

To test how this bias depended on the shape distribution in an extended surround, we entered the biases in a 2 x 4 repeated measures ANOVA with the factors Flanker Condition (60º angle or 140º angle) and Surround Distribution ('centre average', 'flanker average' or , 'homogenous surround'). There was a significant effect of Surround Distribution (*F*(1,7) = 9.09, *p* = 0.000). But, as expected, Flanker Condition showed no main effect (*F*(1,7) = 1.9, *p* = 0.21) or interaction (*F*(3,7) = 1.04, *p* = 0.40). Next, we looked into the main effect of Surround Distribution with a series of planned comparisons (Figure 2.a). The bias in the 'flanker average' and 'homogenous surround' conditions, which differed in heterogeneity but not in mean of the distribution, was equally large *t*(1,7) = -0.19, *p* = 0.85. This shows that the bias was unaffected by the variance in the shape distribution. The bias in the 'center average' condition, however, was smaller compared to the 'flanker average' (*t*(1,7) = -2.81, *p* = 0.026) and 'homogenous surround' conditions (*t*(1,7) = -3.48, *p* = 0.01).

Next, we tested whether the reliability of the shape discriminations depended on the shape distribution in the surround by performing a 2 x 4 repeated measures analysis of variance on the discrimination thresholds in the different Flanker (60º angle or 140º angle) and Surround ('baseline', 'center average', 'flanker average' and 'homogenous surround') conditions (Figure 2.b). But this analysis showed no effect of Flanker (*F*(1,7) = 0.98, *p* = 0.35) or Surround (*F*(1,7) = 1.74, *p* = 0.19) condition on the discrimination thresholds. Thus, there was no effect of the shape distribution on the reliability with which shape discriminations were made.

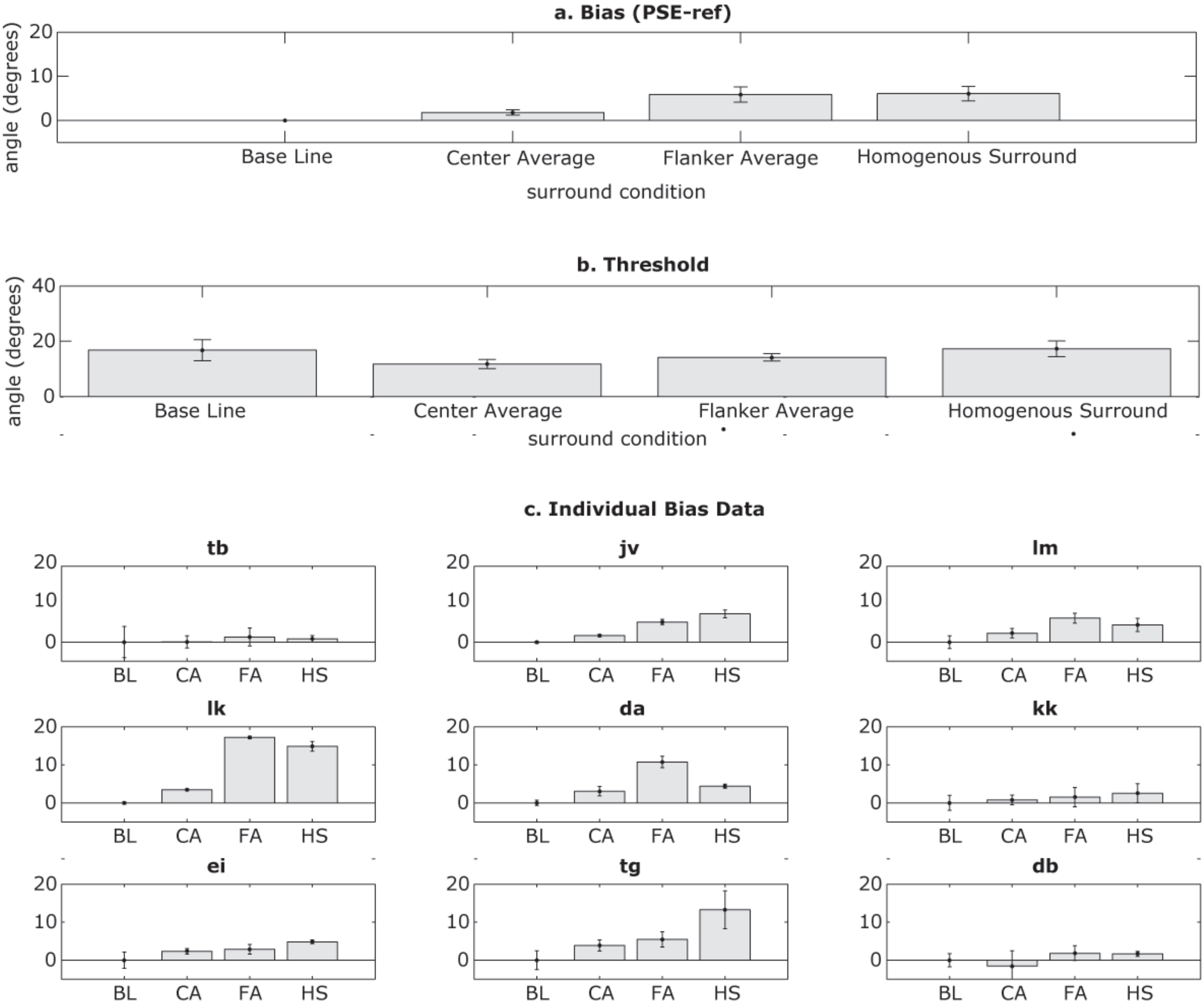


Figure 2. **a.** Mean bias data (PSE – reference angle) pooled over the flanker conditions (60º or 140º flankers). Error bars represent standard errors of the mean. **b.** Discrimination threshold in degrees of angle. **c.** individual bias data, error bars represent the upper and lower confidence interval of the psychometric curve.

Discussion

To investigate how a shape contrast bias depends on the shape distribution over an extended surround, we tested shape perception of a hinged plane, folded over the horizontal axis, with five other hinged planes on each side. The angle difference with the flankers was constant between conditions, but we varied the dihedral angle distribution of the eight outermost surfaces ('extended surround'). We found a shape contrast bias with the flankers, but this bias depended heavily on the angle distribution in the extended surround. A large bias was found when the test shape was different from the mean shape over the entire stimulus, but this bias almost disappeared when the mean shape approached the central shape value while the shape difference with the direct flankers remained constant. Furthermore, the bias, or the reliability with which shape discriminations were made, did not depend on the variance in the shape distribution. The fact that the bias depended on the mean of the shape distribution but not on the variance of the distribution reveals properties of the underlying mechanism. Our results show that shape contrast is caused by a mechanism which takes global properties of a scene into account. We continue to discuss what this means for a proposed MLE and normalization mechanism of shape contrast.

In an MLE mechanism of shape contrast, shape is estimated from weighted linear combination of an absolute and relative shape signal. A bias occurs due to improper cue weighting, where absolute shape is underestimated whereas relative information receives is comparatively overestimated (van Ee et al., 1999). In theory, a global MLE mechanism can explain our results by assuming that relative cues operate not locally, but over a larger area of the visual field. In that case, the difference with the mean of the shape distribution would comprise the relative cue. However in an MLE mechanism, the reliability (variance) of shape signals is the driving force in creating the bias. A hypothesis that is supported by the observation that adding variance to the inducing shapes decreases a contrast bias (van der Kooij & Te Pas, 2009). In a global MLE mechanism, the relative cue is based on the mean of shape distribution and its reliability is affected by the variance of the shape distribution. Yet, our data show that the bias was unaffected by the variance of the shape distribution. Thus, a global MLE mechanism of shape contrast can only explain contextual biases under the unlikely assumption that the variance of the shape distribution does not affect the reliability with which its mean is estimated.

In contrast, a divisive normalization mechanism, where the central shape signal is divided by the shape signal in the surround, does not make predictions on how the bias will depend on the variance of the shape distribution. In a divisive normalization model, a shape contrast bias arises because, for instance, an angle divided by a larger contextual angle, results in a smaller value compared to the same value divided by a smaller contextual angle. Divisive normalization of neural responses to stimuli with a single center and surround stimulus, has been well-established (e.g. Cavanaugh et al., 2002; Sengpiel, Sen, & Blakemore, 1997), but little is known about normalization of shape signals from more complex stimuli. Our psychophysical demonstration of how perceptual biases depend on global properties of a complex stimulus, is of interest to determining the neural architecture of divisive normalization. A likely candidate for the neural implementation of normalization are the horizontal connections between neurons coding for similar features and which can link regions over several millimeters (Series et al., 2003). But normalization may also be accomplished by feedback from higher areas (e.g. Cavanaugh et al., 2002; Zipser, Lamme & Schiller, 1996). First, the speed with which information travels over horizontal connections is not sufficient to account for the time course of normalization (Bair et al., 2003). Second, blocking the transmission of activity through horizontal connections does not fully disrupt normalization (Brown et al., 2003). Because higher areas typically code for more global aspects of a scene, normalization through feedback connections could take into account global properties of a scene. In shape contrast biases are caused by divisive normalization of visual signals, our results are more consistent with a feedback mechanism compared than with a mechanism relying entirely on the horizontal connections between neurons coding for similar features.

Chapter 5

Perception of 3D Slant Out of the Box

Abstract

Evidence for contextual effects in visual perception is widespread. Although this suggests that contextual effects are the result of an inherent property of the visual system, current explanations are limited to the domain in which they occur. In this paper we propose a more general mechanism of global influences on the perception of slant. We review empirical data and evaluate proposed explanations of contextual biases. By assessing a model about 3D slant perception and evaluating more generic mechanisms of contextual modulation, we show that surround suppression of neural responses explains the major phenomena in the empirical data on contextual biases and that contextual biases may be part of a mechanism of grouping and segmentation.

Introduction

A widespread approach to studying visual perception is to isolate an object from the environment and to determine the relation between the information from this object that reaches our eye, and mental processing. This approach disregards important contextual information that is available in the natural environment (Gibson, 1979). In this paper, we address the mechanisms of contextual integration, specific for 3D slant perception. Contextual stimuli provide relative information but can also be used to fill-in missing information. For instance, the visual system receives no visual input where the optic nerve leaves the eye. Yet we usually do not perceive a blind spot because the brain fills in missing information with contextual information. The fact that the visual system uses contextual information is also evident in contextual biases, where perception is biased by contextual stimuli. Phenomena of filling-in and contextual biases show that studying how the visual system integrates spatial information is vital to understanding the mechanisms of visual perception in the natural environment which contains rich contextual information.

Significant advances have been made in the understanding of contextual biases in the perception of 2D orientation. Physiological experiments have revealed interactions in the responses of neurons coding for similar tilt (e.g. Cavanaugh, Bair, & Movshon, 2002; Sengpiel, Sen, & Blakemore, 1997) and neural models of these interactions can predict contextual biases in the perception of orientation (Schwartz, Sejnowski, & Dayan, 2009). But the mechanisms of contextual biases in three-dimensional (3D) perception remain largely unknown (Gillam, Blackburn, & Brooks, 2007). In this paper, we focus on the mechanisms of contextual effects in the perception of 3D slant. To this end we review empirical data and evaluate different explanations of contextual biases. We start with an explanation that has its roots in the modeling of 3D perception from a combination of depth cues but also consider more generic mechanisms that have been proposed to explain contextual biases in the perception of tilt. In the following paragraph, we first summarize the empirical data on slant contrast.

Empirical data on slant contrast

Because slant is not directly available in the retinal image, contextual effects can take place at several levels of processing. Perception of slant is possible because the visual system derives depth relations from depth ‘cues’ such as binocular disparity, motion, perspective, texture and shading. Psychophysical studies have shown that contrast biases in slant perception are not determined by simple motion or disparity cues, which signal relative depth, but are determined by the disparity or motion gradient that specifies slant (Te Pas & Kappers, 2001; Te Pas, Rogers, & Ledgeway, 2000) or relative slant (Gillam et al., 2007; Gillam, Chambers, & Russo, 1988; Gillam & Pianta, 2005). Because disparity and motion gradients offer incomplete information about depth (see Howard & Rogers, 2002 for a review) when evaluated separately, these cues eventually have to be combined. Thus, contextual biases could either arise in the processing of individual disparity and motion gradients or at a processing stage where these slant signals have already been combined. In a recent study (van der Kooij & Te Pas, 2009a), we showed that contrast and assimilation biases originate at a different level of slant processing. A slant contrast bias occurs before cue combination, within the individual cue systems, whereas an assimilation bias originates at a stage of slant processing where individual slant signals have been combined.

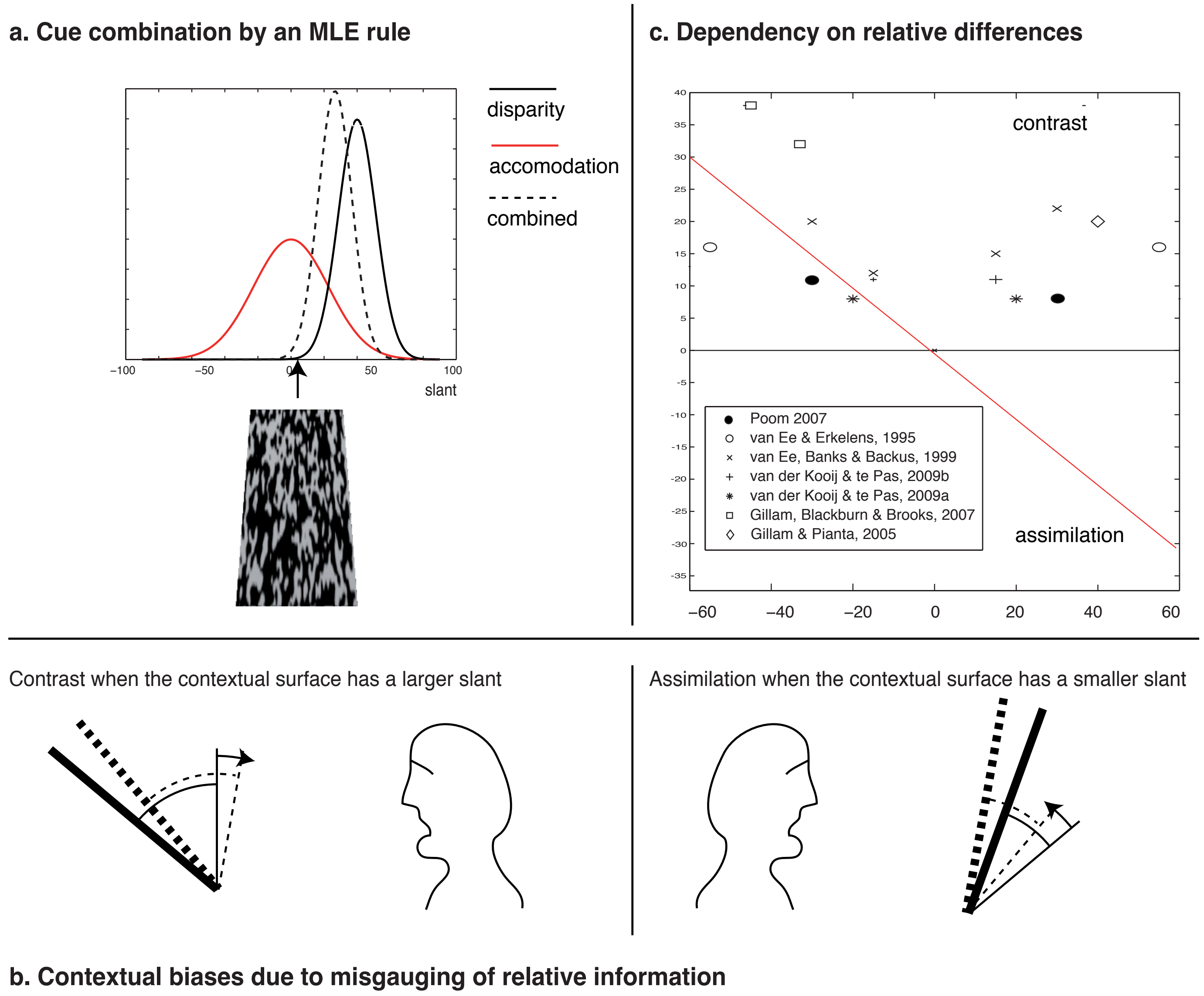
Now it has been determined in which slant signals contextual biases originate, we proceed to determine how biases depend on the information in these slant signals. Contextual biases do not depend on whether the contextual surfaces have a larger or smaller slant compared to the central surface (van der Kooij & Te Pas, 2009a, 2009b). This shows that the bias depends on the difference with contextual surfaces and not on the absolute slant of these surfaces. The bias that contextual stimuli invoke, depends not only on the relative information they provide but also on the reliability of slant information. When slant information is reliable, a contrast bias is observed that decreases with degraded reliability of contextual slant. When the central slant information is unreliable, on the other hand, an assimilation instead of contrast bias is observed (van der Kooij & Te Pas, 2009b). In contrast with the dependency of contextual *biases* on the reliability of slant information, we found no evidence that the *reliability* of the central shape percept depends on the presence of contextual stimuli.

Despite the vast amount of data on slant contrast, the mechanisms producing contextual biases in perception of 3D shape remain obscure. In this paper, we shine light on the mechanisms of contextual biases in slant perception by evaluating different explanations of contextual effects. We first discuss a computational explanation that has its roots in the modeling of 3D perception from a combination of depth cues, but also consider a more generic mechanism that has been developed to explain contextual biases in 2D tilt perception.

Maximum Likelihood Estimation

Because contextual stimuli provide relative information about slant, contextual biases can be described as part of mechanism that integrates different types of information to increase the reliability of perceptual judgments. A maximum likelihood estimation (MLE) model of slant perception describes how the visual system overcomes uncertainty in individual slant signals by combining these signals. In such a model, combination occurs by taking a weighted average of slant signals. This means that when a slant signal, say binocular disparity, is unreliable, it will have less influence on perception and other slant signals, such as accommodation and perspective, will have more effect. MLE models successfully describe how 3D slant is estimated from a combination of depth cues (e.g. (Knill, 2003; Landy, Maloney, Johnston, & Johnston, 1995). In theory, MLE models can incorporate contextual biases by treating contextual surfaces as relative cues to slant, which are also weighted. Typically, the visual system is rather insensitive to scale and shear within a surface (Gillam et al., 1988; Mitchison & Westheimer, 1984; Shipley & Hyson, 1972; Stevens & Brookes, 1988; van Ee & Erkelens, 1996) but is more sensitive to differences between surfaces (Gillam, Flagg, & Finlay, 1984; van Ee & Erkelens, 1996). Therefore, relative information is the more reliable cue and, together with underestimation of absolute slant, this is thought to cause a contrast bias (van Ee, Banks, & Backus, 1999; van Ee & Erkelens, 1996). Straightforward as this rationale may seem, it is unclear how the visual system combines absolute and relative information to slant because the two types of information are qualitatively different. An absolute cue to slant specifies the orientation of a surface whereas a relative cue specifies the slant *relative to a reference*. Hence, averaging of the absolute and relative cue can only produce veridical perception after cues have been gauged to an appropriate standard (i.e. the mean slant in the stimulus). Gauging relative information to an inappropriate standard can cause biased perception.

One MLE model of slant perception made an attempt to formalize how absolute and relative information are combined (van Ee et al., 1999). This model attributes a slant contrast bias to the situation where stereograms are presented on a flat computer screen, which introduces cue conflict: a single cue (say binocular disparity) signals a slanted surface whereas other cues signal a flat surface. Combination of slant by disparity and other cues results in an estimate of absolute slant that is lower than disparity-defined slant (Figure 1.a). A ‘contrast’ bias occurs because a relative disparity signal is gauged to an underestimation of the slant of the contextual surface. However, this MLE model predicts contrast only when the contextual surface has a larger slant compared to the test surface. When the contextual surface has a smaller slant, an assimilation bias is predicted (Figure 1.b). This prediction does not hold, as equivalent biases have been found when the contextual surface has a larger or smaller slant compared to the test surface (van der Kooij & Te Pas, 2009a, 2009b); Figure 1.c). So, gauging relative cues to an underestimation of slant predicts a *shift* in slant estimates, not the contrast enhancement found in psychophysical experiments.


a. Cue combination by an MLE rule

c. Dependency on relative differences

b. Contextual biases due to misgauging of relative information

Figure 1. a) Cue combination by an MLE rule. Combination of disparity and accommodation cues causes underestimation of slant. **b)** Contextual biases when relative slant is gauged to an underestimation of contextual slant. **c)** Contextual biases reported in the literature plotted as a function of relative slant. From the studies cited in this review, we chose the biases that were found in the perception of disparity-defined slant around a horizontal axis, with flanking contextual stimuli that were separated by a small gap. These biases are compared to how contextual biases depend on relative differences in the Slant Estimation Model (red line, van Ee et al., 1999).

Furthermore, in an MLE model, visual signals are combined to increase reliability of perceptual judgments. If information from contextual stimuli is integrated in the slant estimate, this should result in more reliable slant judgments. In the literature, we found no evidence that reliability of slant judgments increase in the presence of contextual surfaces. Furthermore, in an MLE model, slant signals are combined according to their reliability and the more reliable signal should have the most effect. Again, this prediction is not supported by empirical data: when the reliability of the contextual signal is reduced, the bias they invoke is indeed reduced (van der Kooij & Te Pas, 2009b; van Ee et al., 1999) but when the central slant signal is degraded and the contextual signal is the more reliable one, an assimilation instead of contrast bias is observed (van der Kooij & Te Pas, 2009b). This shift from contrast to assimilation is not predicted by a simple MLE model. The only way in which an MLE model can predict the shift is by assuming that, when slant information is unreliable, the absolute slant signal from the test and contextual surface are integrated as if they were caused by the same visual object. This assumption is unlikely as observers were well able to perceive separate surfaces when surface slant was degraded with added noise (van der Kooij & Te Pas, 2009b).

To summarize, although MLE models have been very successful in explaining how slant is estimated from a combination of depth cues, they do not easily explain contextual biases in these slant estimates. More promising explanations may be found outside the domain of 3D perception.

Surround suppression

Contextual stimuli affect not only perception of a central stimulus but also neural responses to this stimulus. In of ‘surround suppression’, the neural response to a stimulus presented on its receptive field is reduced when a similar stimulus is presented to in the surround of its receptive field. In the early psychophysical literature on slant contrast the bias has been linked to the neural phenomenon of surround suppression (Anstis & Howard, 1978; Brookes & Stevens, 1989; Mitchison & Westheimer, 1984; Schumer & Ganz, 1979) but at that time it was unclear how *suppression* of a neural response would cause a perceptual *bias*. A recent computational model (Schwartz et al., 2009) shows how

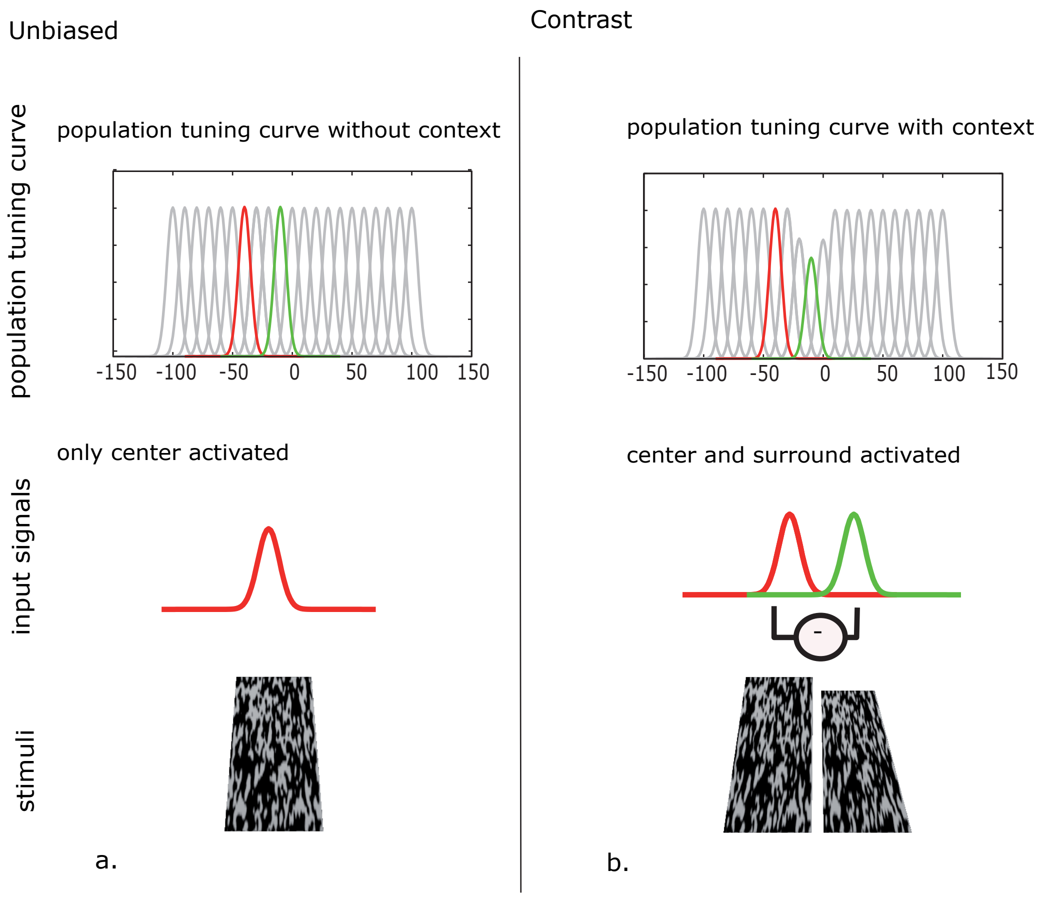


Figure 2. Suppression of responses to similar stimuli causes a contrast bias in the population tuning curve.

suppression of responses to similar stimuli causes a contrast bias in the population tuning curve from which a visual property is decoded (Schwartz et al., 2009; Figure 2).
If surround suppression causes a contrast bias, the bias will be largest when the contextual surface stimulates the suppressive surround of the receptive field of the neuron coding for the central stimulus and will wear off when the contextual slant falls off the suppressive surround of the receptive field (Cavanaugh et al., 2002; Figure 3.a). To test this prediction, we modeled the neural tuning curves for slant with a Mexican Hat Function (difference of Gaussians or DoG model) (Sceniak, Ringach, Hawken, & Shapley, 1999), encompassing both the excitatory center and the suppressive surround of the receptive field. The suppressive surround was modeled as twice the size of the excitatory center and the only free parameters in the model were the width of the tuning curve and a scaling parameter. The dependency of contrast biases on slant difference between the test and contextual stimulus was well described by a difference of the tuning curves coding the test and contextual slant, each having a standard deviation of 22 degrees of angle ($R^2 = 47.7$, $p = 0.002$; Figure 3.c). This shows that a contrast bias in slant perception is well described by a mechanism operating on surround suppression. Furthermore, the observation that slant contrast can be described with the difference of two Mexican Hats is consistent with a recent observation by Mareschal, Morgen and Solomon (2010) that the dependency of the tilt illusion on cortical separation between the center and surround can be described with a Mexican Hat function.

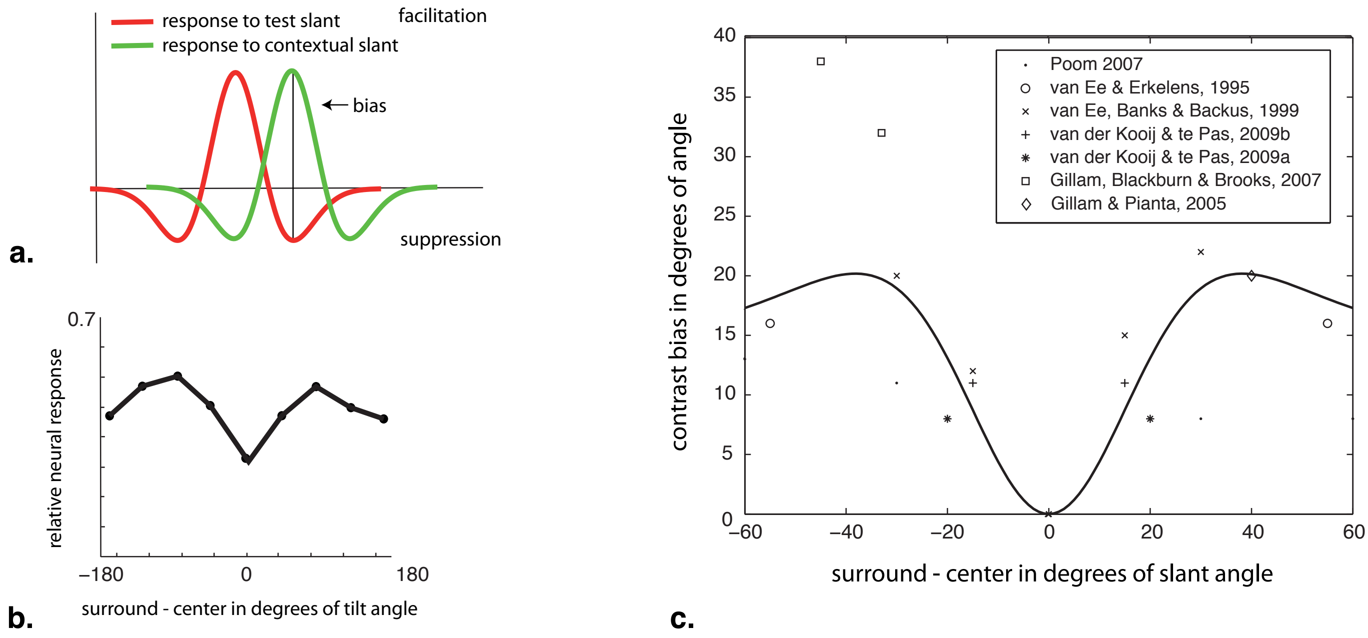


Figure 3. a) Surround suppression is determined by the difference of the tuning curve coding for the test and contextual surface. **b)** Example electrophysiological data from V1 showing response suppression for neurons associated with the center stimulus as a function of the discrepancy between center and surround orientations. Data replotted from Cavanaugh et al. (2002). **c)** Slant contrast biases in degrees of angle found in different experiments plotted as function of the slant difference with the contextual surface. The solid line represents the fit of a difference of Mexican Hats (with standard deviation and scaling as a free parameters) for the central and contextual stimulus.

The shift in bias direction with reliability of visual information can also be seen in surround suppression. Recent physiological experiments show that neurons coding for similar orientation do not only inhibit but may also facilitate each other, depending on the strength of visual input (Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Sceniak et al., 1999). A finding that has been confirmed with human fMRI data (Tajima et al., 2010). Just as surround suppression causes a contrast bias in the population tuning curve, surround facilitation causes an assimilation bias in the population tuning curve. The fact that assimilation occurs also when the shape of central and contextual surfaces is defined by a different depth cue, can be explained indirectly by the phenomenon of surround suppression. Slant from cue combination is processed in higher visual areas (Tsutsui, 2002; Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005) where receptive fields are typically large (Dumoulin & Wandell, 2008). When receptive field size increases, at some point the contextual stimuli no longer stimulate the suppressive surround but fall onto the facilitatory center, which would cause assimilation.

As shown above, the phenomenon of surround suppression shows parallels with contextual biases in slant perception, which suggests a causal link between the phenomena. However, physiological data on surround suppression versus facilitation of neural responses to tilt have not directly been linked to perceptual biases and the link between suppression versus facilitation a contrast versus assimilation has not been demonstrated empirically. Novel fMRI techniques allowing for population receptive field measurements such as the study by Tajima and colleagues (2010) allow for the simultaneous measurement of surround suppression and perceptual biases and can provide more detailed answers. Also, surround facilitation has only demonstrated with lowered contrast and the question whether added noise can induce surround facilitation still has to be answered. Moreover, the link between the neural and perceptual phenomenon does not answer the question on the functionality of shape contrast. Answers may be found outside the domain of 3D perception as contextual biases in slant and tilt perception depend in a similar way on relative angle differences (Figure 4).

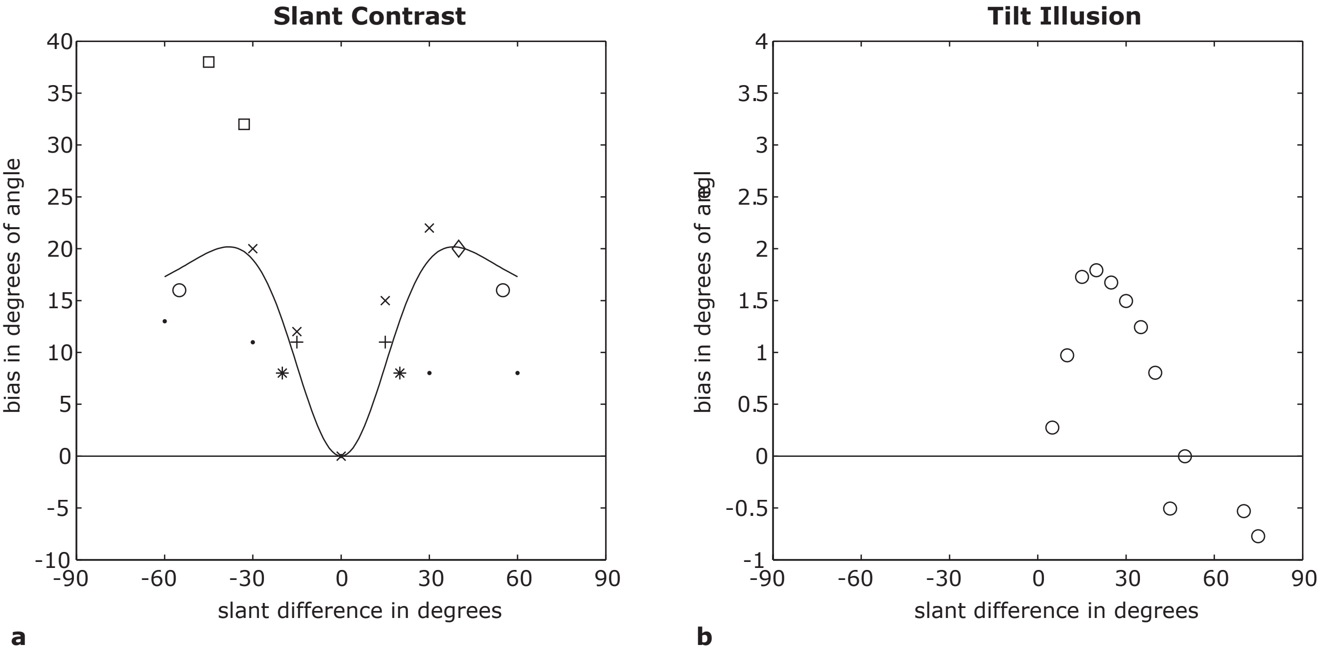


Figure 4. There are similarities in the dependency of contextual biases in slant (**a**) and tilt (**b**) perception on relative angle differences. Tilt data are replotted from (Westheimer, 1990) and averaged over subjects.

Grouping & segmentation

A recent divisive normalization model by Schwartz, Hsu and Dayan (2009) explains both tilt contrast and assimilation from a mechanism of grouping and segmentation. When two tilt signals are held to originate from a common world source, the visual system normalizes the response to these stimuli by the total tilt signal. This way, the common factor in a region is averaged out and the response to similar stimuli is reduced. On the neural level, reduction of responses to similar stimuli could be achieved by surround suppression (Carandini, Heeger, & Movshon, 1997; Geisler & Albrecht, 1992; Heeger, 1992; Wilson & Humanski, 1993). Counter intuitively, a contrast bias occurs when two tilt signals group because the response to similar stimuli is reduced. When two tilt signals are held to originate from a different world source, they are not included in the normalization pool, and no contrast bias occurs. In this case neurons may even facilitate each other, increasing the neural response (and salience) at the border between groups and creating an assimilation bias in the population tuning curve. In this model, the visual system determines whether two signals originate from a common source by taking advantage of the statistical correlation structure of natural images. For instance, the response of two neurons coding for similar slant is often correlated whereas the response of two neurons coding for very different slant will be less correlated. This rationale explains why in the tilt illusion assimilation instead of contrast is observed for large angle differences. In the slant domain such large angle differences have not been measured.

The model by Schwartz and colleagues (2009), in its current form does not explain why a contrast bias changes to assimilation when information is unreliable. Nor does it explain why assimilation occurs when central and contextual slant are defined by a different depth cue. However, theoretically, assimilation with unreliable information is consistent with the idea that contrast occurs between correlated stimuli whereas assimilation occurs between uncorrelated stimuli. Noisy signals give less evidence of correlation and therefore may not induce surround suppression.

The hypothesis that contextual biases are the result of a mechanism of grouping and segmentation cannot be tested with existing data but brings forth fascinating novel questions. In general, the effect of grouping on a slant contrast bias in the spatial domain has received little attention. Only the effect of spatial proximity on a slant contrast bias is well known. As predicted by Schwartz's model, bias size decreases with spatial separation (Gillam & Pianta, 2005; Poom, Olsson, & Borjesson, 2007). However, spatial proximity of the test and contextual surface not only affects grouping of the two surfaces, but also changes the disparity content of the stimulus. The effect of grouping by feature similarity on a sequential slant contrast bias (aftereffect) has received some attention. An aftereffect in relative depth perception has been shown to be specific for surfaces that are defined by the same texture (Blaser & Domini, 2002) or color (Domini, Blaser, & Cicerone, 2000). These data are consistent with the hypothesis that slant contrast biases occur between temporally grouped elements but do not allow for conclusions in the spatial domain.

Conclusion

Contextual effects are widespread in visual perception and may reflect a generic property of the visual perception in natural scenes where rich contextual information is available. Contextual effects in slant perception are poorly explained by an MLE model of 3D perception where slant is estimated from a combination of absolute and relative information about slant. Instead, contextual biases in slant perception are better explained outside the domain of 3D perception. Physiological data on surround suppression and facilitation of neural responses to tilt are well in accordance with how perceptual biases in slant perception depend on relative differences and the reliability of visual information. Phenomena of surround suppression and facilitation can be considered part of mechanism of grouping and segmentation as shown by a recent computational model of the tilt perception. These observations bring forth important novel questions on contextual effects in slant perception.

The relation between contextual biases and neural surround suppression and facilitation will have to be more firmly determined by simultaneous measurement of neural responses and perceptual behavior. In addition to existing MLE models, this research may reveal novel ways in which the visual system deals with uncertainty in three-dimensional information. The visual system may not only integrate visual information to increase reliability, as in an MLE model, but may also actively ignore redundant information when information is clear.

The hypothesis that contextual biases are the result of a grouping and segmentation process brings forth an exiting novel field of research. By measuring contextual biases of slanted elements in a complex stimulus it can be assessed whether slant contrast is indeed reduced at the border between groups. Next, it will be of interest to study whether contextual biases in slant perception are affected by segmentation by other visual properties, such as texture. This way, we will come closer to understanding the mechanisms of slant perception in natural scenes.

Chapter 6

Surface Boundaries Do Not Constrain a Depth Aftereffect

Under review at Vision Research

Abstract

We investigated whether a depth aftereffect is better explained in a surface-based organization of depth representation (Nakayama & Shimojo, 1992), where depth is represented in conjunction with surface features, rather than a coordinate-based organization, where depth is represented in conjunction with spatial position. Observers adapted to a stereogram depicting a surface with reference plane and were tested with a surface with either the same contour or position as the adaptation surface. The aftereffect did not depend on test surface position or contour. Thus, a depth aftereffect can be caused by a mechanism that does not depend on grouping by surface contour or position.

Introduction

Aftereffects have become a popular paradigm to investigate the organizational principles of perception. Many organizational principles are feasible, of which the most intuitive a coordinate system as used successfully used in physics. Different coordinate systems are possible. The external world can be described in a three-dimensional coordinate system where positions are defined on a horizontal, vertical and depth axis, which is called a spatiotopic coordinate system. But the visual information that the brain receives, is projected onto a two-dimensional retinal surface. When an observer moves the eyes, the retinotopic coordinate system translates, whereas the spatiotopic coordinate system does not, creating inconsistency between the two coordinate systems. On the other hand, when a motionless observer fixates, the coordinate systems are aligned. Experiments investigating which of these coordinate systems is used in visual perception show that adaptation is not always position dependent.

Depth aftereffects do not depend on the continuous stimulation of retinal coordinates as depth aftereffects have been found when, during adaptation, observers made eye-movements over are a depth stimulus (Noest, 2006; Ryan & Gillam, 1993) or with oscillating disparity in the adaptation stimulus (Berends, 2001). Recently, aftereffects have been reported at positions with no overlap to the adapted position in any coordinate system (Taya, Sato, & Nakamizo, 2005). Although these aftereffects appear free from any coordinate system, Taya and colleagues explain this position independence by adaptation of mechanisms where global features are processed and receptive fields are large. Thus, depth is represented on a position in some coordinate system, but on such a large spatial scale that the visual system was insensitive to the change in physical position.

All the aforementioned studies implicitly rely on the assumption that the visual system retrieves the depth of positions on the visual field and that depth representation is coordinate-based. Remarkably, few studies have tested the fundamental assumption whether indeed the visual system uses a metric structure of coordinates to represent depth (Thaler & Goodale, 2010). Alternatives can be proposed which possibly better explain the contradictory findings obtained to date. In this paper, we answer the question whether the phenomenology of an aftereffect is better explained in a coordinate-based model of depth representation or in a surface-based, more abstract, organization of depth representation that we describe below (Nakayama & Shimojo, 1992).

There are reasons to be skeptical towards a purely coordinate-based system for depth processing. In the natural images, densely textured regions are infrequent, and information about depth can be sparse (Nakayama & Shimojo, 1992). Nakayama and Shimojo (1992) showed that in sparse stereograms observers perceive an arrangement of surfaces in depth instead of depth coordinates interpolated from the contours (Figure 1a). More recently, it has been shown that the perception of structure-from-motion is influenced by surface symmetry in a way that cannot be explained by surface interpolation (Treder & Meulenbroek, 2010). Therefore, depth perception might be surface-based rather than coordinate-based (Nakayam & Shimojo, 1992; Treder & Meulenbroek, 2010) which means that the visual system directly associates the retinal image with a stored surface representation instead of reconstructing the depth of individual positions on the visual field. In this theory, surfaces are represented through an associative linkage of visual features. This way the visual system can recall the whole pattern of features when, due to for example transparency, an image contains only a subset of these features. Recently, psychophysical evidence has been found for such linkage of feature representation (Blaser & Domini, 2002; Domini, Blaser, & Cicerone, 2000). Depth aftereffects can be specific for surface features such as color (Domini, Blaser, & Cicerone, 2000) and texture (Blaser & Domini, 2002). This means that an aftereffect occurs between two surfaces that are defined by the same color or texture but not between two surfaces that are defined by a different color or texture. Thus, instead of retrieving the visual properties of spatial coordinates and updating this coordinate system over time, the visual system constantly forms associative linkages between surface features and compares novel images to these linkages.

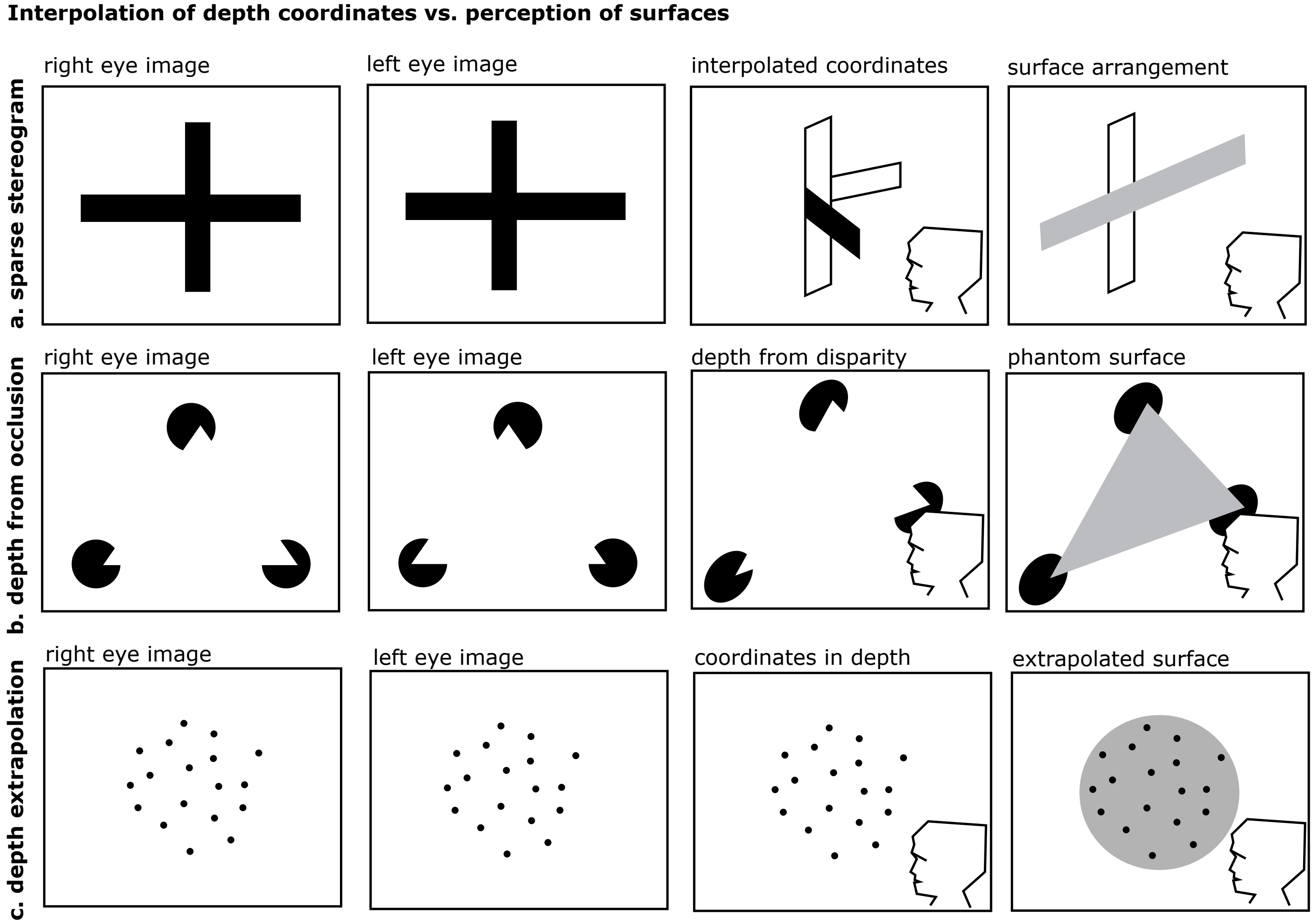


Figure 1. Examples of coordinate based versus surface based depth perception, in the left two columns stereograms, prepared for cross fusion are presented. In the third column the percept predicted from a coordinate-based model of depth perception is presented at in the outmost right column the predictions from a surface based model of depth perception is presented. Gray areas represent the filled-in depth coordinates. (**a**). Example taken from Nakayama and Shimojo, 1992. Most observers perceive a horizontal bar in front of a vertical bar instead of the figure predicted from depth interpolation. (**b**) In depth perception from occlusion most observers perceive a sharply defined phantom plane. (**c)** In a random dot stimulus, depth is extrapolated from the texture elements until the illusory contours of the transparent surface.

Surface contours might constitute a key element of surface-based depth representation, even more so than color or texture. Depth can be perceived from occluding contours alone (Figure 1b), taking advantage of the fact that distant surfaces are occluded by nearer surfaces to a different extent in the two eyes (i.e. Anderson & Nakayama, 1994). Interestingly, the depth region that is ‘filled-in’ from monocular occlusion regions is bound by a sharp ‘knife like’ occluding contour. This shows that the visual system can recover depth regions with lacking positional information and, moreover, that it assigns depth to a region *within* an occluding contour. In fact, Ryan and Gillam (1993) explain the observation that adaptation to two discrete lines in depth resulted in a position independent aftereffect by arguing that observers adapted to the disparity gradient *between* the two lines, instead of to relative disparity per se. But contours not only provide information to mechanisms of filling-in (surface interpolation), they also constrain processes of surface extrapolation. In transparency, for example, information about depth has to be extrapolated (spread out) from the texture elements on the surface *until* its boundaries. Random dot stimuli, where high luminance dots on a hypothetical surface float in black space, are an instance of such transparent surfaces (Figure 1.c). When surface boundaries are ambiguous because the depth step between surface and background is ill-defined, depth may spread beyond the hypothetical surface. This way, there might be position overlap in the neural representation of the adaptation and test surface even if there was no physical overlap, causing ‘position independent’ aftereffects (Taya et al., 2005).

The general hypothesis that depth is represented in conjunction with surface contour leads to three hypotheses on a depth aftereffect, which we test in three experiments. In all experiments, observers adapt to a square surface in front of a reference plane and perform a depth discrimination task on a sequentially presented reference plane and test surface, of which we vary the contour and position contingency with the adaptation surface. In experiment 1 we test whether a depth aftereffect is contour contingent, not position contingent. In experiment 2 we test whether the visual system uses contour information to differentiate between surfaces when it adapts to two different surfaces in depth. And in a final experiment we test whether reports of position invariant aftereffects are due to weak surface boundaries, allowing depth to extrapolate freely.

General Methods

Stimuli & apparatus

Observers adapted to a small square surface in front of a larger reference plane defined by a pacman texture (Figure 2.a). To avoid a priori exclusion of large receptive field adaptation as an explanation of position dependency, stimulus sizes were chosen to ensure that the entire stimulus fell within the receptive fields of neurons in higher visual cortex such as area V5 and LOC. In this areas, single cell recordings in primates have shown receptive field measurements range from 8 to 16 degrees of visual angle (Felleman & van Essen, 1991; Maunsell, 1987). The adaptation surface measured 1.5 degrees of visual angle in width and height whereas the reference plane was 7.5 degrees of visual angle wide and 5 degrees of visual angle high. A nonius symbol was presented 3 degrees of visual angle above the center of the adaptation surface so observers could monitor their fixation. Because observers fixated, overlap in position relative to the eye coincided with overlap in position relative to the head. The individual pacmans on the reference plane had a diameter of 1.2 degrees of visual angle. Adaptation surface and reference plane pacman texture luminance was constant at 66 cd/m^2.

The reference plane was at fixation depth and thus contained zero disparity. The adaptation stood four centimeters in front of the reference plane and its distance was defined by binocular disparity in all experiments. Depth of the test surface, on the other hand, was varied around a comparison depth of

2.3 cm (12 arcminutes disparity) in a range from 0.4 cm (12 arcminutes disparity) to 3.8 cm (22.2 arcminutes disparity) in front of the reference plane. In experiment 1 and 2, test surface depth was signaled by disparity but in experiment 3 by monocular occlusion. Stereograms were presented on a mirror set up with two Philips 40 x 30 cm Brilliance 202P7 CRT monitors, set to a resolution of 1920 by 1440 pixels. Viewing distance was 48 cm and a chinrest was used to maintain head position.

Depth discrimination task

Observers performed a depth discrimination task, while asked to maintain fixation on the nonius symbol. No instructions on what section of the stimulus to attend to were given. Observers indicated whether the test surface stood further from the reference plane compared to a top view comparison stimulus. In this comparison stimulus, a wide bar (15 degrees of visual angle) represented the reference plane and a smaller bar (10 degrees of visual angle) represented the test surface (Figure 2.c). We first calibrated this top view for each observer. To this end, observers matched the top view to stereograms depicting the test surface at seven different depths, varied around comparison depth. Each depth was replicated four times for each of the experimental conditions and the top view comparison stimulus for the following discrimination tasks was the average depth setting for the comparison depth at 2.3 cm. Depth settings were used not only to calibrate the comparison stimulus, but also as a test of stereo acuity. Two observers whose depth settings showed no relation to simulated depth were excluded from further participation in the experiment.

To allow for stereo vision, the test stimulus (surface and reference plane) was presented for 1 second, follow by a black screen inter-stimulus interval of 500 ms, followed by comparison stimulus. Now observers indicated whether the test surface stood further from or closer to the reference plane compared to the comparison stimulus, using the up and down arrows on the keyboard. After an inter-trial interval of 1 second, the next trial started. In the adaptation condition, we followed a top up paradigm where adaptation blocks started with 60 seconds adaptation and each trial started with 7 seconds adaptation.

Staircase procedure

We measured the point of subjective equality (PSE) for depth perception using a staircase procedure. For each condition, four staircases were measured: two starting with a stimulus depicting a surface at a larger depth from the reference plane compared to the comparison stimulus (‘down staircase’) and two starting with a stimulus depicting a surface at a smaller depth from the reference plane (‘up staircase’). Pilot experiments showed that observers were able to maintain fixation and stereo vision up to thirty minutes maximum. This allowed us to measure 6 reversals of each staircase in the adaptation condition. Therefore, all staircases were terminated after 6 reversals and we obtained the PSE by averaging the simulated depth from the reference plane on the last 4 switch trials from all 4 staircases. We subtract simulated comparison depth from these PSE’s to obtain a measure of bias in depth perception in the different conditions.

Procedure

The experiment was divided over three sessions, measured on different days and each taking about one hour. Within in an experimental block we measured one ‘up’ and one ‘down’ staircase for each position and contour condition. In the adaptation condition of experiment 1, this took up to 25 minutes. During experimental blocks, observers were allowed a short break each five minutes to relax fixation, but they maintained head position in the chin rest.

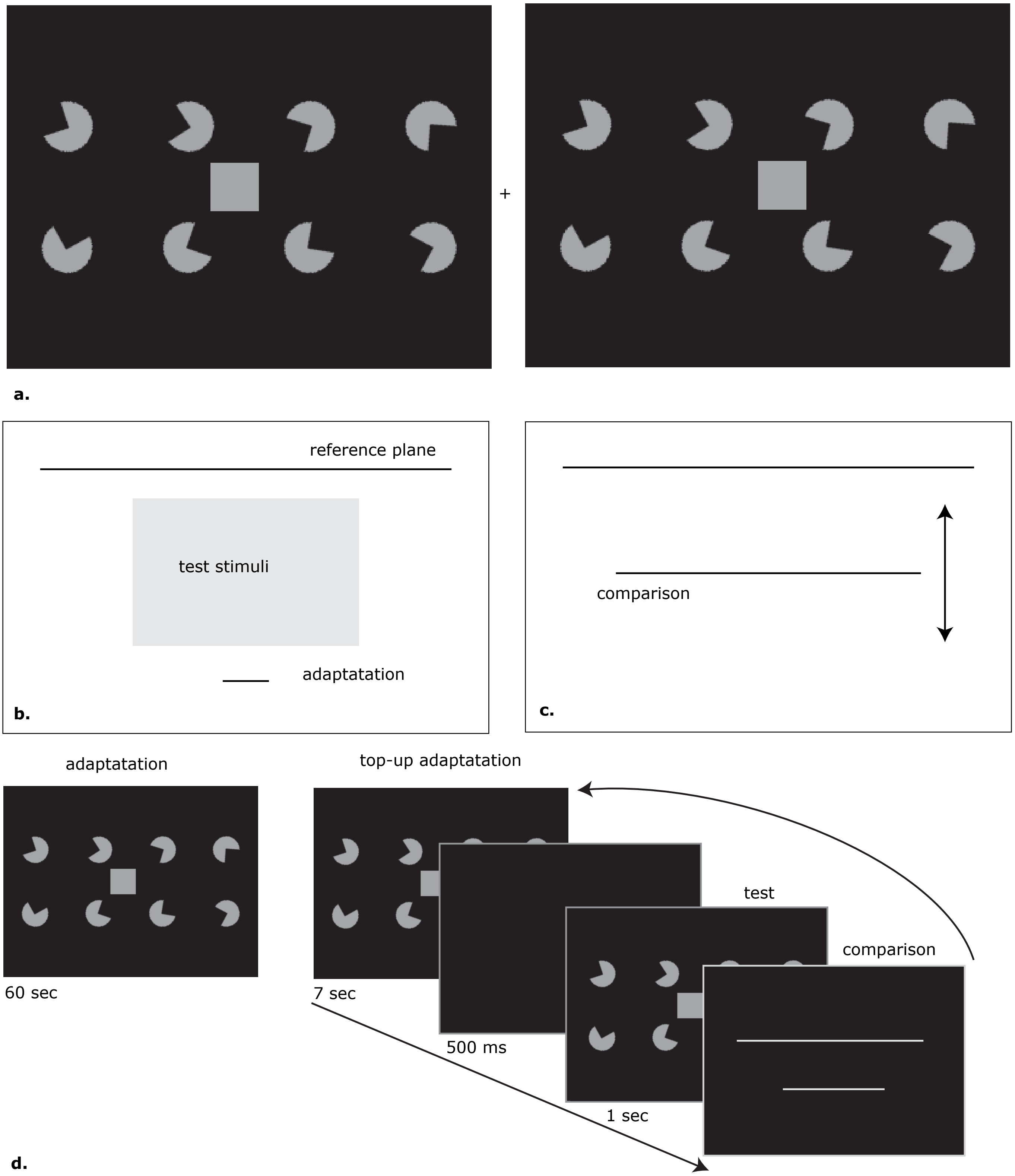


Figure 2. a. Stereogram of the adaptation stimulus, prepared for cross fusion. **b.** Top view cartoon of the depth layout of reference plane, adaptation and test stimuli. **c.** Top view, which the observers could match in the calibration session and to which they compared simulated depth in the depth discrimination task. **d.** Trial sequence. Anaglyph demonstrations of the stimuli can be viewed on: www.katinkavanderkooij.com/thesis/chapter6/abstractChapter6.html

Experiment 1

To investigate whether depth is represented in conjunction with surface contour, we tested the hypothesis that a depth aftereffect is contour contingent, not position contingent. This implies that an aftereffect can be found independent of measured position but that adaptation to a square surface will cause a larger bias in the perception of a sequentially viewed square surface compared to a triangular surface.

Methods

Eight observers, including one of the authors (KK) participated. The other participants were Psychology students at Utrecht University. They had normal or corrected to normal vision and participated for course credit or a financial reward of 6 euro per hour. To induce relative depth adaptation, observers adapted to a square surface in front of a reference plane, maintaining fixation on a nonius symbol 3 degrees of visual angle above the centre of the adaptation surface. To test whether a depth aftereffect is contour contingent, we tested a depth aftereffect on either the same square surface or on a triangular surface in front of the same reference plane (Figure 3). Surface area was constant between square and triangular surfaces at 4.34 cm^2 and the triangular surface overlapped with 85 % of the square surface. To assess whether a depth aftereffect is position contingent, the test surface was presented either at the same central position as the adaptation surface or displaced 2.5 degrees to the left.

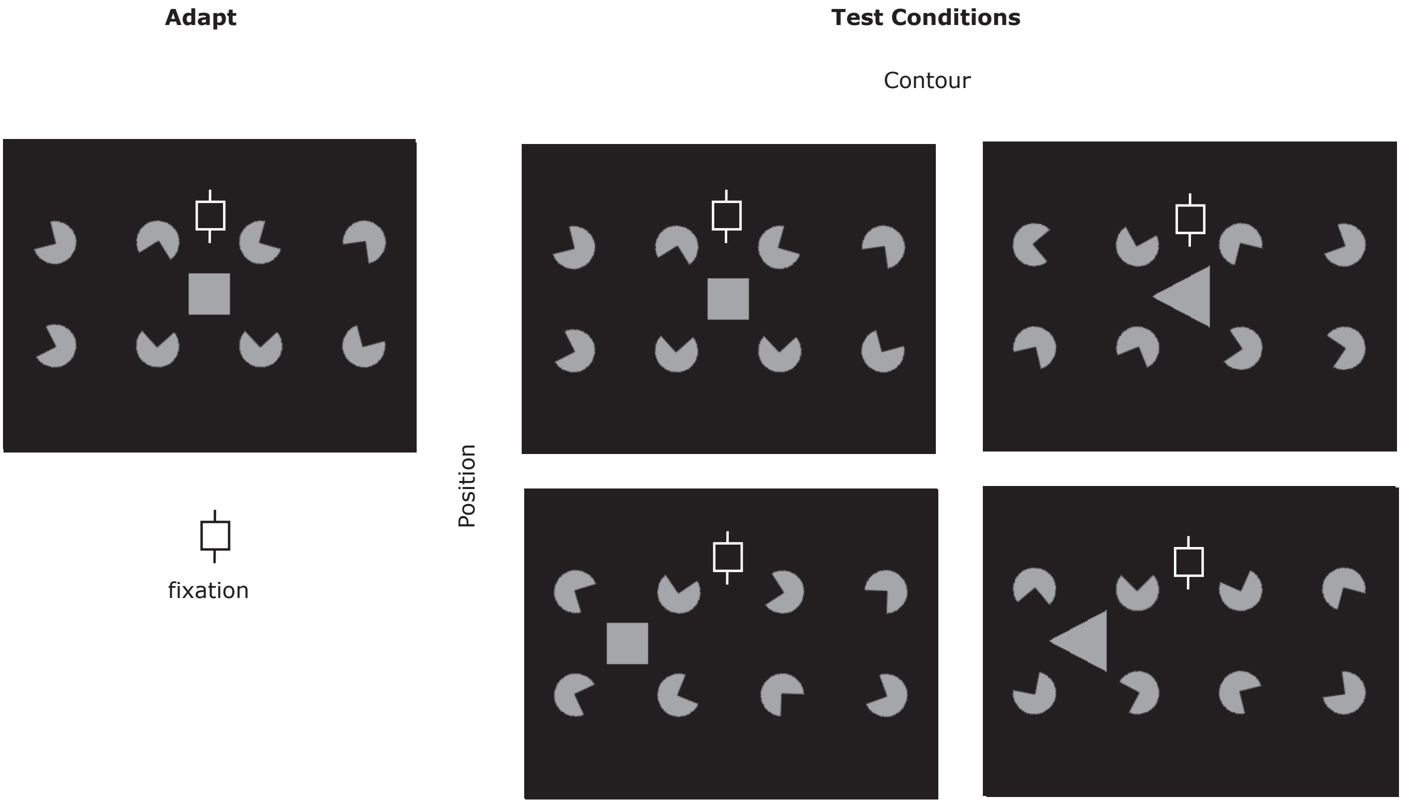


Figure 3. Test surface position and contour conditions in experiment 1. Observers maintain fixation on the nonius symbol while viewing the adaptation and test stimuli.

Results

To test whether adaptation to a square surface caused a negative bias in the depth perception of a sequentially viewed surface, and to assess whether this effect depended on position and contour overlap, we entered the bias data (comparison depth – PSE) in a repeated measures analysis with the factors Adaptation (baseline, adaptation), Contour (square, triangle) and Position (same, different). There was a main effect of Adaptation ($F(1,7) = 43.44$, $p = 0.00$), reflecting a negative aftereffect where adaptation to a large relative depth caused the relative depth in a sequentially viewed test stimulus to be underestimated (Figure 4.b). But this main effect did not interact with Contour ($F(1,7) = 1.85$, $p = 0.22$) or Position ($F(1,7) = 0.48$, $p = 0.83$). This shows that the aftereffect did not depend on overlap in contour or position between the adaptation and test stimulus.

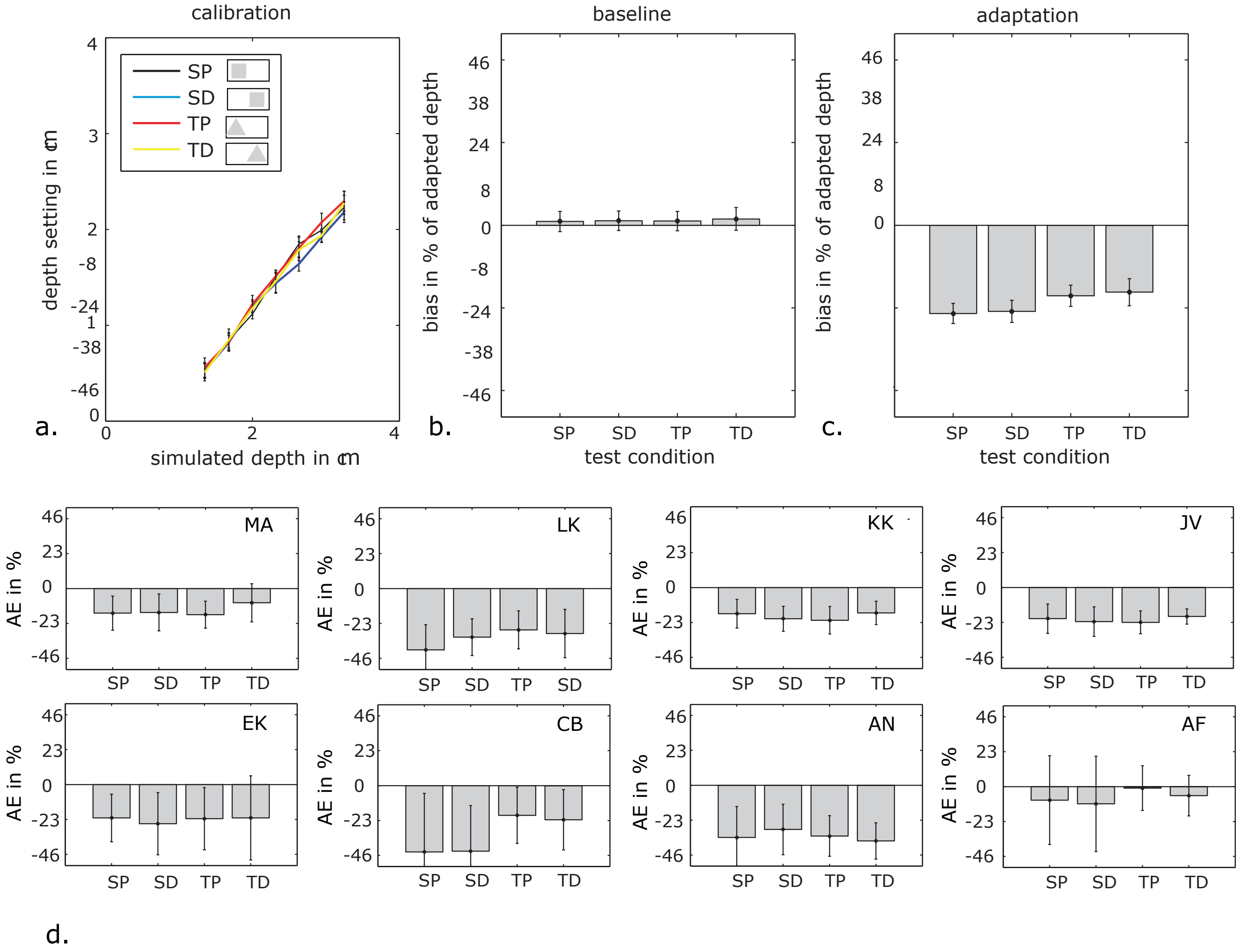


Figure **4. a.** Average depth setting and standard deviation for the simulated depths in the different test conditions (SP = square, same position, SD = square, different position, TP = triangle, same position, TD = triangle, different position. **b & c** Average bias (comparison - PSE) in percentage of adapted depth in the different contour and position conditions. **b.** Baseline. **c.** Bias after adaptation. Data are averaged over 8 subjects and error bars represent standard errors of the mean. **d**. Individual aftereffects (AE) in percentage of adapted depth (PSE baseline – PSE adaptation). Errors bars represent standard deviations of the mean.

Discussion

We measured an aftereffect in a depth discrimination task, which did not depend on overlap in position or contour between adaptation and test stimulus. This is evidence that depth was not represented in conjunction with contour or position. However, the visual system might encode surface features in the depth representation, but only when it needs to differentiate between multiple surfaces. In experiment 1, the visual system was stimulated with a single relative depth between surface and reference plane and did not need to differentiate between depth stimuli.

Experiment 2

In experiment 2, we test the hypothesis that when the visual system adapts to multiple surfaces at different depths, it encodes not only their relative depth but also surface features by which it can differentiate between them. Observers adapted to not one, but two surfaces and a reference plane. One surface was positioned 2.5 degrees to the left of fixation and placed in front of the reference plane, whereas the other surface was positioned 2.5 degrees to the right of fixation and placed at an equal depth behind the reference plane. As the reference plane was at fixation depth, the two surfaces had a relative disparity with an opposite sign, and the average disparity signal in the stimulus was zero. Thus, an aftereffect of relative depth can only be found if the adapted mechanism is sensitive to the conjunction of depth and position or to the conjunction of depth and contour. If no aftereffect occurs, this is evidence that the visual system adapts to the pooled disparity over a larger region, as would be predicted by large receptive field adaptation such as proposed in the literature (Taya et al., 2005).

Methods

Eight observers, of whom only KK participated in experiment 1, adapted to both the triangle and square test surface used in experiment 1, one placed 2.5 degrees to the left of the center and the other place 2.5 degrees to the right. In one adaptation condition, the square was placed in front of the reference plane and the triangle behind and in the other adaptation condition, this order was reversed. To test for a contour or position contingent depth aftereffect, two test conditions were created. Hereby, we took advantage of the fact that adaptation to the surface that was placed in front or behind the reference plane predicts opposite aftereffects. As only adaptation to the surface that was placed in front would predict an aftereffect as in experiment 1, the properties of the test surface are compared to this surface. In a position condition, observers were tested with a different surface at the position of the adaptation surface that was placed in front of the reference plane. But in a ‘contour’ condition, observers were tested with the same surface contour, but at the position of the surface that was placed behind the reference plane in the adaptation stimulus. Thus, position and contour contingency would predict aftereffects in opposite directions.

To test whether observers were sensitive to disparity sign (in front / behind), they matched a top view to both the situation where the surface was placed in front or behind the reference plane. Trials where the surface was placed in front or behind were intermixed. The subsequent depth discrimination task was performed on the stimuli where the test surface was placed in front of the reference plane.

Results

To check whether there was a difference between the condition where observers adapted to the square in front and the triangle behind or the reverse configuration, data were first entered into a 2 x 2 x 2 repeated measures ANOVA with the factors Configuration (square in front, triangle in front), Adaptation (baseline, adaptation) and Test Condition (location, position). As there was no difference between the adaptation conditions where the square or the triangle was presented in front ($F(1,7) = 0.79$, $p = 0.44$), data were pooled from these conditions. Contrary to findings in experiment 1, a repeated measures analysis with the factors Adaptation (baseline, adaptation) and Test (position, contour) showed no effect of adaptation, $F(1,7) = 0.66$, $p = 0.44$, nor main effect of Test ($F(1,7)$, $= 2.95$, $p = 0.13$ or interaction of Test and Adaptation ($F(1,7) = 0.012$, $p = 0.95$) (Figure 5).

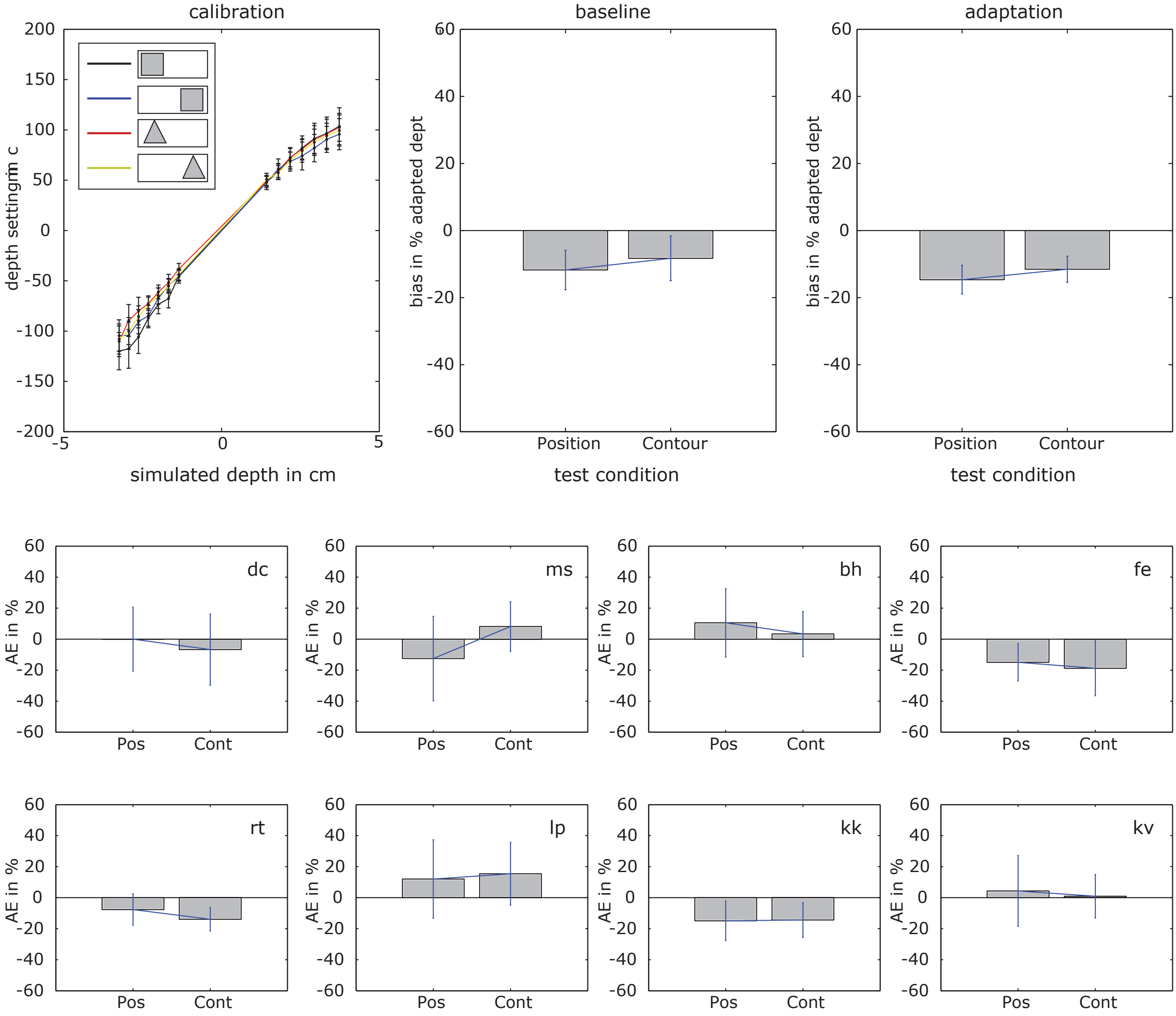


Figure 5 a. Depth settings as a function of simulated depth in the calibration session for the different test surfaces (square left, square right, triangle left, triangle right. b. Bias in percentage of adapted depth in the two test conditions (position, contour). c. Bias in percentage of adapted depth after adaptation in the two test conditions (position, contour). Error bars represent standard errors from the mean. d. Individual aftereffects (AE) in percentage of adapted depth (PSE baseline – PSE adaptation). Here, error bars represent standard deviations from the mean.

Discussion
In contrast with evidence that aftereffects can be highly specific, either for position (i.e. Mitchell & Baker, 1973) or surface features (e.g. Domini et al., 2000), we found that when adapting to two surfaces and a reference plane, the visual system adapts to their pooled depth rather than to their individual distances to the reference plane. Thus, we again find no evidence that depth was represented in conjunction with surface contour or position.

Moreover, these results allow us to reject two alternative hypotheses on why we find no position or contour contingent aftereffect. First, measured position-independency was not due to eye movements which brought the adaptation and test stimulus to the same retinal position. This could have been achieved only by reversing the depth of the left and right test stimulus: an impossible task. If observers adapted to zero depth, because wandering eye-movements spread the two depth signals over the retinal image, the left and right section of the stimulus, spanning an area of five degrees visual angle, would have had to be sampled equally. Yet spontaneous eye-movements during a prolonged fixation task have been found to remain within a one-degree area (Knapen & van Ee, 2006) and we consider this explanation unlikely. Second, the aftereffects were not due to adaptation to the reference plane or to overall depth compression of the stimulus. As the perception of a reference plane can influence perceived depth of another stimulus (Glennerster & McKee, 1999; He & Ooi, 2000), adaptation to the reference plane could have caused a bias in the perception of the test figure. However, the same reference plane as in experiment 1 was used, and overall depth was greater than in experiment 1, so both adaptation to the reference plane or overall depth compression predict an aftereffect, which is not what we found.

If, on the other hand, observers adapted to the pooled depth signal because the surfaces extrapolated beyond their luminance contour, strength of the surface boundary would affect the area where an aftereffect can be found. That is, a surface interpolates and extrapolates within surface boundaries (Nishina, Yazdanbakhsh, Watanabe, & Kawato, 2007). In experiment 3 we answer the question whether the strength of surface contours affects the area where an aftereffect can be found.

Experiment 3

In experiment 3 we test the hypothesis that reports of position independent aftereffects are due to weak surface boundaries, allowing depth to extrapolate freely. To isolate the effect of surface boundaries, the aftereffect was tested on a phantom surface that was defined by the texture on a reference plane. Thus, there was no overlap in local disparities between the adaptation and test surface. In experiment 1 and 2, the test surface floated in empty space where the depth step between surface and reference plane did not have to occur at the surface contour. In experiment 3, we strengthened surface contours. To this end, we placed the test surface on top of the texture of the reference plane, creating occlusion cues. In occlusion, the depth step has to occur *at* the surface contour. Now, if depth spreads through interpolation and extrapolation, constrained by surface contours, we expect that a negative aftereffect can be found on the phantom surface. Moreover, we expect that in this stimulus, depth extrapolation stops at the surface contour and that a depth aftereffect only occurs when there is overlap between the adaptation and test surface.

Methods
The eight observers that also participated in experiment 1, adapted to the same disparity defined square and reference plane as in experiment 1. To test whether an aftereffect in relative depth perception can occur on an interpolated surface, observers were tested with a larger phantom surface with a width and height of 2.5 degrees visual angle defined by the pacmans on the reference plane

(Figure 6). Depth of the phantom surface was specified at its boundaries, which gave rise to monocular occlusion regions on the texture of the reference plane. To assess whether the depth aftereffect is still position independent when surface boundaries are disambiguated by occlusion, two position conditions were created for the test surface. In a position-contingent condition, the test surface was presented centrally, as was the adaptation surface, but in a position-invariant condition it was displaced 2.5 degrees to the left or to the right.

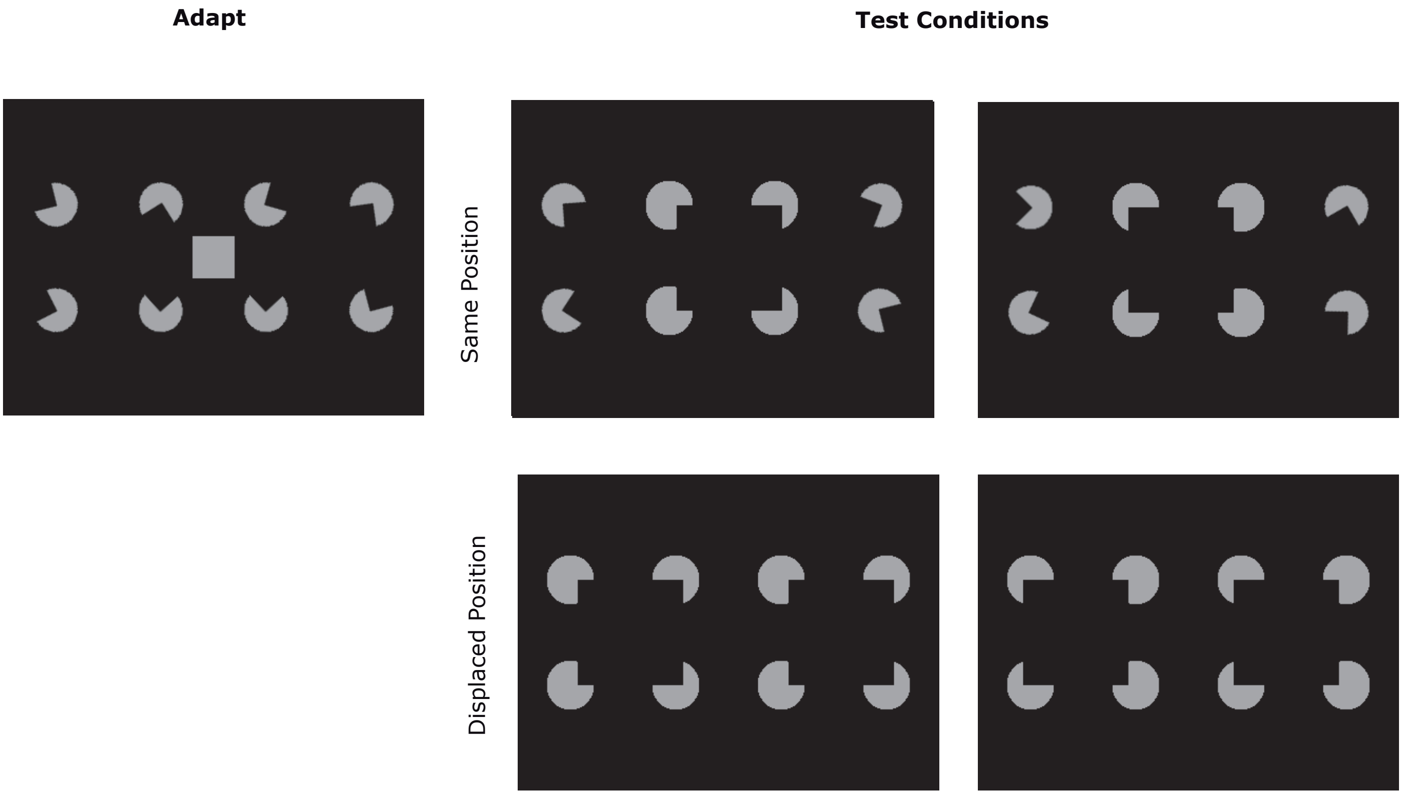


Figure 6. Adaptation and test surface conditions in experiment 3. The illustrations of the test surface are stereograms, prepared for cross fusion.

Results

To test whether adaptation to a disparity defined square caused a negative aftereffect in the relative depth perception of a phantom surface, and to check how this effect depended on position of the test stimulus, bias data (Fig. 7) were entered in a repeated measures ANOVA with the factors Adaptation (baseline, adaptation) and Position (same, different). As in experiment 1, there was a significant main effect of Adaptation, reflecting a negative aftereffect ($F(1,7) = 6.23$, $p = 0.047$). But there was no interaction with Position ($F(1,7) = 2.87$, $p = 0.14$), which shows that the size of this aftereffect was independent of measured position.

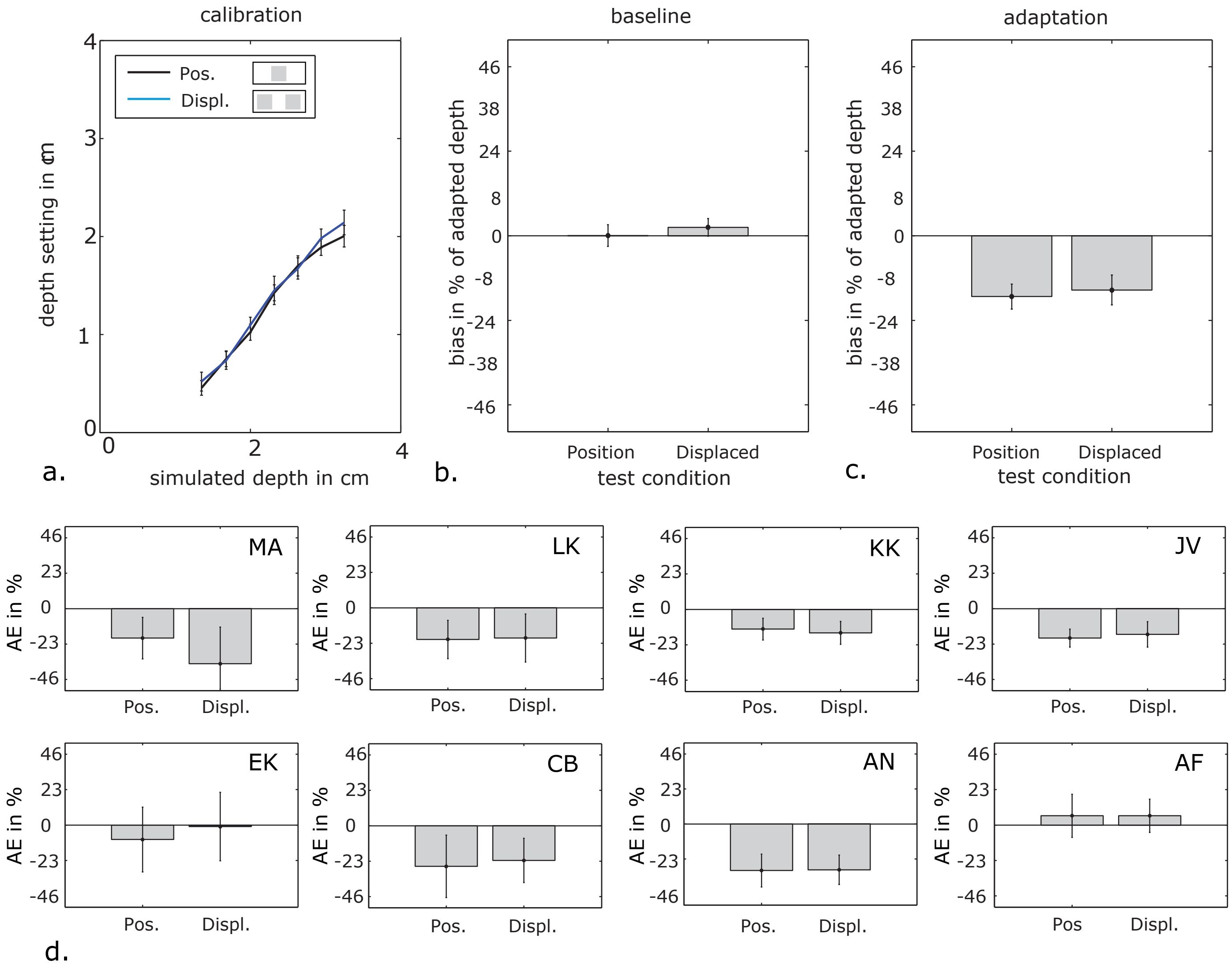


Figure 7. a. Depth settings for the different simulated depths in the calibration session for the two test conditions (same position, displaced position). **b.** Bias comparison – PSE) without adaptation in percentage of adapted depth in the two test conditions. **c.** Bias in percentage of adapted depth after adaptation in the two test conditions. Error bars represent standard errors of the mean. **d.** Individual aftereffects in percentage of adapted depth (PSE baseline – PSE adaptation). Error bars represent standard deviations of the mean.

Discussion

We found cross adaptation between a small disparity-defined square surface and a larger phantom surface where depth was specified at the boundaries, giving rise to monocular occlusion regions on a reference plane. As there was no overlap in local depth signal between the adaptation and test stimulus, this shows that an aftereffect can occur on an interpolated surface. But the aftereffect did not depend on the position of the test surface. Thus, even with strong surface boundaries the aftereffect did not depend on the position of the test surface. The aftereffect was smaller in size, compared to the aftereffect in experiment 1. We explain this by the fact that the phantom surface had a much larger area compared to the disparity defined test stimulus. Therefore the relative intensity of the adaptation and test stimulus signal was different between the two experiments. In the motion domain, indirect evidence exists that relative intensity of the adaptation and test stimulus affects the size of an aftereffect, as the largest motion aftereffect is found when the adaptation stimulus has high luminance contrast but the test stimulus low luminance contrast (Ishihara, 1999; Keck, Palella, & Pantle, 1976). In conclusion, an aftereffect in relative depth can be found on an interpolated surface, but surface boundaries do not constrain the area where this aftereffect can be found.

General Discussion

Recently, depth aftereffects have been reported at different positions on the visual field than the adapted position (Taya et al., 2005). This suggests that the adapted depth representation was not represented in a coordinate system. At the same time, depth aftereffects can be specific for surface features such as texture and color (Blaser & Domini, 2002; Domini et al., 2000), which suggests that these properties *are* included in the adapted depth representation. In this paper, we investigated whether depth is represented in an abstract surface-based organization (Nakayama & Shimojo, 1992) rather than in a coordinate system. In a surface-based organization of depth perception, surface representations are formed by an associative linkage of surface features, such as color, texture, depth and contour (Nakayama & Shimojo, 1992). Thus, adaptation to a surface in depth would result in an aftereffect on a surface with common features but not on a surface defined by different features. For instance, a square surface can invoke an aftereffect on a square surface at a different spatial position but not on a triangular surface (a contour contingent aftereffect).

We measured a negative (contrasting) aftereffect in relative depth perception that was independent of surface position or contour (square or triangle). Even when the visual system was stimulated with two surfaces at a different depth, position and with a different contour, it adapted to the pooled depth of the two surfaces, although observers were well able to perceive two separate surfaces. Whereas occlusion constrains depth extrapolation, it did not constrain the area where the aftereffect was found. Thus position independence of the depth aftereffect cannot be attributed to depth extrapolation beyond the contours of the physical surface (Taya et al., 2005). With these results, we provide the first experimental evidence that a surface based organization of depth representation does not explain recent findings of position independent depth aftereffects.

Taya et al. (2005) tested a similar hypothesis –whether a depth aftereffect is shape specific- but their results are inconclusive. In their study, observers adapted to random dots on a transparent circular surface and were tested with random dots on an annulus. They conclude that the aftereffect was unspecific for shape (circle / annulus), but the depth from the annulus could have been perceptually filled-in to compromise the inner circle, creating overlap with the adaptation stimulus. This was also noted by Taya and colleagues who performed a second experiment where observers adapted to a rectangular surface and were tested with the same surface at a different position, still finding an aftereffect. However, in this condition, there was again overlap in surface shape between adaptation and test stimulus. Thus, based on the Taya experiments, the hypothesis that a depth aftereffect is contour specific cannot be rejected. But together with our results, the hypothesis that depth is represented in conjunction with surface contour can be rejected as an explanation for a position independent aftereffect.

Nevertheless, one might ask whether depth is in fact represented in conjunction with surface contour but on a much larger spatial scale than we measured. Psychophysical evidence has shown that adaptation to three-dimensional stimuli can affect the mapping of disparity to slant (Berends, Liu & Schor, 2005) or the processing of higher-order disparity signals (Domini, Adams, & Banks, 2001; Ryan & Gillam, 1993). Receptive fields are typically large in cortical areas where higher-order signals are processed and spatial resolution in these areas is held to be low (Felleman & Van Essen, 1991; Maunsell & Newsome, 1987; Van Essen, Newsome, & Maunsell, 1984). But if we filter our occlusion-defined stimuli with a Gaussian modeled to the receptive fields of neurons in higher visual cortex (Dumoulin & Wandell, 2008), no depth is perceived. Therefore the mechanisms that processed these stimuli must have been sensitive to information on a relatively small spatial scale.

The issue of spatial scale is more relevant to explain the fact that the aftereffect did not depend on position of the test surface. The position of the adaptation and test surface might have been represented on such a low spatial resolution, that their representations overlapped even though their physical positions did not. This argument has been brought forth by Taya and colleagues (2005) and can also be found as an explanation for position independence in the motion domain.

Finally, one might ask whether the aftereffect was in fact position-dependent, with eye movements bringing the adaptation and test stimulus to the same retinal position. Our data make this explanation unlikely. When observers adapted to two surfaces with a horizontal separation, they adapted to the average depth of the two surfaces. The only pattern of eye movements that can invoke this result is equal sampling of the left and right section of the stimulus. This would have required eye movements to wander over an area over five times larger than the area over which eye movements usually wander during adaptation (Knapen & van Ee, 2006).

To conclude, we report an aftereffect in relative depth perception that was not specific for spatial position or surface contour. This result contrasts with reports of feature specific aftereffects (e.g. McCollough,1965), which have resulted in the frequent use of aftereffects as a psychophysical tool to study the sensitivity to conjunctions of features (Blaser & Domini, 2004). Now, aftereffects are being used as a psychophysical tool even though the neural mechanism and functionality of adaptation remain largely unknown (Kohn, 2007). Yet, definite conclusions can only be drawn from these experiments when the mechanisms of adaptation are known. In determining the underlying mechanism, the specificity of aftereffects is important (Kohn, 2007). Our results show that depth aftereffects might be caused by a less specific mechanism than previously thought.

Chapter 7

Summary and Discussion

The basic problem of studying the perception of shape in complex scenes is that the general lay out and geometry of the entire scene can easily affect the subjective appearance of shape. Striking examples of the effect of contextual stimuli on shape perception are contrast biases where the shape difference with contextual stimuli is perceptually enhanced. Such effects are the consequence of the fact that visual perception is often based on relative rather than absolute measures. Our brain does not register the shape of an object, it registers the shape of an object *relative to the rest of the scene*. Yet, shape perception has mainly been studied in isolation. These studies have focused on the question how information from different depth cues is combined into a coherent shape percept. In this thesis we have gained insight in the *mechanisms* that combine all contributing cues in a scene to a consistent whole.

We have addressed the mechanisms by which the visual system integrates contextual information by studying two well-known biases in shape perception: a contrast bias and an assimilation bias. To be able to link our psychophysical data to current models of shape perception, we study contextual biases in the perception of a simple hinged plane folded around the horizontal axis. A hinged plane offers the advantage that its shape is defined by two slants around the horizontal axis, allowing for a comparison of the data with both the shape and slant perception literature. Here, we summarize the main findings in this thesis and the impact that these findings have for the understanding of the mechanisms of shape perception in complex scenes.

Because perception of shape is usually derived from a combination of depth cues, such as disparity and motion, contextual effects can take place at different levels of shape processing: at a 'within-cue' level, where information from individual cues is processed, or at a 'combined-cue' level, where information from different cues is combined and processed. The first goal in this thesis was to find out at which level of processing shape contrast effects originate. This will provide us with new insight in the mechanisms behind the perception of shape, and allow us to refine the search for the *mechanism* that produces contextual effects in shape perception. In the literature, it has been proposed that contextual effects take place at a 'within-cue' level of shape perception (van Ee et al., 1999). In Chapter 2, we show that contextual effects not only take place at a within cue level, but also at a 'combined-cue' level. We created a 'combined-cue' stimulus where the shape of central and contextual surfaces was defined by a different depth cue (disparity or motion), and compared shape perception in this condition to perception of a 'within-cue' stimulus where all shapes were defined by the same depth cue (again, disparity or motion). In the within-cue condition, the contextual surfaces invoked a contrast bias, but in the combined-cue conditions, where shape interactions had to take place at a 'combined-cue' level of shape perception, the contextual surfaces invoked a small assimilation bias.

The assimilation bias in the combined-cue condition can be explained by the fact that shape from cue-combination is typically processed in higher visual cortex (Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005), where receptive fields are large (Dumoulin & Wandell, 2008). These large receptive fields integrate shape information over a larger area, which may have encompassed both the central and contextual stimuli, with an assimilation bias as a result. The mechanism that produces shape contrast, on the other hand, has to be sought at a 'within-cue' level of shape perception. Interestingly, in the within-cue condition we did not only find contrast biases. Some observers also showed an assimilation bias in the condition where all shapes were defined by motion, and where interactions could occur on a within-cue level of shape perception. For these observers, shape estimates based on motion were highly variable. This suggests that contextual effects are dependent on the reliability of shape signals.

The relation between contextual biases and reliability of shape signals is interesting because the visual system might integrate contextual information to increase the reliability of shape judgments. This idea can be formalized in a statistical model of maximum likelihood estimation (MLE). In an MLE model of shape perception, shape signals are averaged weighted by their reliability. Such models have been successful in describing how shape is estimated from a combination of depth cues and have been proposed for the integration of direct and contextual information about shape as well (van Ee et al., 1999). In an MLE model of contextual effects in shape perception, integration of relative information from contextual stimuli causes a contrast bias. This bias will be largest when contextual information is reliable. We test the feasibility of an MLE mechanism of contextual integration by determining how contextual effects in shape perception depend on the reliability of direct and contextual signals to shape. We varied the reliability of information about shape by adding spatial noise to the central shape, the contextual shapes, or both, and compared shape perception between conditions. The main result is that when the central shape signal was reliable, a contrast bias occurred whereas an assimilation bias was observed when the central shape signal was degraded with added noise. This change in bias direction is not predicted by an MLE mechanism, which predicts an increased contrast bias when the contextual stimuli are the more reliable one. The change in bias direction is also interesting because it shows that contrast and assimilation biases might arise from a common mechanism.

Because current MLE mechanisms of shape perception from a combination of shape cues do not explain why a contrast bias changes to an assimilation bias when shape information is unreliable, we conclude that a better explanation of contextual effects in shape perception is needed. To find the mechanisms of contextual effects in shape perception, we further narrow down the type of information that drives the bias.

The distinction between processing of local and global information is an important one because different neural areas have been associated with the processing of local and global information. In Chapter 4, we address the issue whether contextual effects in shape perception are due to local or global mechanism. Relative differences exist both on a local level between adjacent surfaces and on a global level, where the relative shape difference between a surface and the global shape of the stimulus is determined. In the literature, shape contrast effects have been attributed to local differences between surfaces (Gillam, Blackburn & Grooves, 2007) but have also been attributed to global shape differences between contextual stimuli and a central surface (Curran & Johnston, 1994). In Chapter 4, we investigated whether a contrast bias in shape perception is determined by local shape differences with adjacent surfaces or by a difference with global properties of the stimulus. We presented a central shape with not one but five surfaces of varying shape on each side. This created a distribution of shape signals in the stimulus of which we could vary the mean shape and the variance. A contrast bias in perception of the central shape did not depend on the local shape difference with adjacent surfaces but was determined by the difference with the global (average) shape in the stimulus. However, the variance in the shape distribution had no effect on the bias. As in Chapter 3, these results are not consistent with an MLE mechanism of shape perception. Because contextual biases depend on relative shape difference with the global stimulus, but not on the reliability or variance of these shape differences, we sought to explain contrast bisaes by a global mechanism that makes the visual system sensitive to relative differences.

Contrast biases in shape perception might be caused by the neural mechanism that makes the visual system especially sensitive to relative differences. Sensitivity to relative differences is achieved by inhibitory interactions between neurons coding for similar stimuli. This way, responses to similar stimuli are suppressed, but relative *differences* remain. Inhibitory interactions between neurons coding for shape have been suggested as an explanation for contrast biases in the early literature on shape contrast (Anstis & Howard, 1978; Brookes & Stevens, 1989; Mitchison & Westheimer, 1984; Schumer & Ganz, 1979). But over the recent years, this explanation of shape contrast has received little attention, possibly because a clear concept of how these inhibitory interactions would produce a contrast bias was lacking. At the same time, significant advances have been made in the neurophysiological understanding of inhibitory interactions between neurons, also termed ‘surround suppression’.

Physiological experiments have shown that responses of neurons coding for similar properties are not only suppressed but may also be facilitated, depending on the reliability of visual signals. When visual stimulation is unreliable, neurons tend to facilitate instead of suppress the responses of neurons coding for similar stimuli (Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Sceniak, Ringach, Hawken, & Shapley, 1999); Sceniak et al., 1999), a finding that has been confirmed with human fMRI data (Tajima et al., 2010). Because surround suppression has the character of divisive normalization where the common factor in a stimulus is averaged out (Heeger, 1992), computational models describe surround suppression in a divisive normalization framework. A recent computational model of divisive normalization (Schwartz et al., 2009) links neural suppression and facilitation to contextual effects in the perception of 2D tilt. Surround suppression causes a contrast shift in the population tuning-curve from which tilt is decoded, whereas surround facilitation causes an assimilation shift in the population tuning-curve. In short, when visual signals are reliable, neurons suppress each other, with a contrast bias as a result and when visual signals are unreliable, neurons facilitate each other, with an assimilation bias as a result. This is exactly what we observed in Chapter 4. Thus, a divisive normalization model is better able to explain how a contrast bias depends on the reliability of shape signals than an MLE mechanism.

Furthermore, if contrast biases in shape perception are due to inhibitory interactions between neurons coding for similar properties, the size of a contrast bias should depend on the receptive field properties of the neuron coding for the central shape. Suppression, and consequently a contrast bias, is largest when a contextual stimulus falls exactly onto the suppressive non-classical receptive field and will wear off when the stimulus is so far in feature space that it falls off the receptive field. To test this prediction, we performed a meta-analysis on slant contrast biases that have been reported in the literature (Chapter 5). As predicted by a divisive normalization mechanism, the dependency of the size of a slant contrast bias on relative slant differences in a stimulus was well described by the difference of the neural tuning curve coding for the central stimulus and the tuning curve coding for the contextual stimulus.

Because a divisive normalization mechanism explains how contextual biases depend on the reliability of shape signals and on the relative differences in a stimulus, we conclude that contextual biases in shape perception are caused by a neural mechanism of divisive normalization that makes the visual system especially sensitive to relative differences. An important question that remains to be answered is what the functionality of divisive normalization is. When the common factor in a stimulus is averaged out, this reduces redundancy. Therefore, surround suppression has been attributed to a mechanism that reduces redundancy in visual input allowing for faster processing of the remaining information (Schwartz et al., 2007; Schwartz & Simoncelli, 2001). Surround facilitation, on the other hand integrates visual signals and increases signal strength when information is unreliable. Hence, contrast and assimilation biases could reflect a balance between the visual system ignoring redundant data and obtaining the best possible signal from unreliable data.

Meaningful redundancy reduction can be achieved by grouping and segmentation. Therefore, it has been proposed that contrast effects in visual perception occur especially within grouped regions (Schwartz et al., 2009). A prediction that is supported by empirical data that show that surround suppression in the detection of horizontal offset between two vertically aligned lines (Vernier detection) only occurs when the target and surround do not segment into two independent visual objects. (Herzog, Schmonsees, & Fahle, 2003). In a final chapter (Chapter 6), we addressed the issue whether contextual effects in shape perception are related to the grouping dynamics in a stimulus. Depth separation is a strong grouping cue and grouping and relative shape differences maybe be confounded in a standard shape contrast stimulus. To overcome this confound, we took advantage of the fact that contextual biases also occur in the temporal domain, where prolonged exposure to a stimulus causes a contrast bias in a sequentially viewed stimulus (aftereffect). We tested how an aftereffect in relative depth perception depends on two types of grouping parameters: spatial position and surface contour. Two points in time at the same spatial location often belong to the same visual object and the visual system might group by spatial position. Because visual objects may move, the contours of an object may be the more effective grouping cue. To test the effect of grouping by spatial position or by surface contour, subjects adapted to a surface in depth and were tested with a surface with either the same position or contour as the adaptation surface. We found a contrasting aftereffect of the adaptation stimulus, but this aftereffect did not depend on whether the test stimulus had the same contour or position as the adaptation stimulus. With this result, we find no evidence for a relation between grouping dynamics in a stimulus and contextual effects in perception.

To conclude, contrast biases are the result of the mechanism that makes the visual system especially sensitive to relative differences. This mechanism does not only allow us to register the shape of an object *relative to the rest of the scene*, it also reduces redundancy in visual information. Assimilation biases, on the other hand, are associated with global processing of shape. This means that to be able to understand how our perception relates to the information in a stimulus, it does not suffice to understand how the visual system combines all contributing cues in a scene to a consistent whole. The mechanisms by which the visual system *ignores* certain types of information are just as important.

Nederlandse Samenvatting

Hoofdstuk 1: Inleiding

Visuele context beïnvloedt waarneming van 3D vorm. Een bal temidden van grotere ballen ziet er bijvoorbeeld kleiner uit dan dezelfde bal temidden van kleinere ballen. Ondanks het feit dat zulke contextuele effecten op vormwaarneming in veel psychofysische experimenten zijn aangetoond, wordt vormwaarneming gewoonlijk in isolatie bestudeerd. Een dergelijke aanpak gaat voorbij aan het feit dat context stimuli waardevolle informatie over vorm kunnen bieden. Context stimuli bieden relatieve vorm informatie die bijvoorbeeld specificeert dat een van de ballen groter is dan de andere ballen. In dit proefschrift bestuderen wij de mechanismen van vormwaarneming in complexe scènes waarin contextuele stimuli waarneming kunnen beïnvloeden. Contextuele effecten zoals in het voorbeeld met de ballen geven een kijkje in de mechanismen die vorm en context informatie integreren. Twee typen contextuele effecten treden op. In een *contrast bias* wordt het verschil tussen vorm en context perceptueel uitvergroot en in een *assimilatie bias* wordt het verschil tussen vorm en context juist perceptueel verkleind. In een serie psychofysische experimenten onderzoeken we hoe deze biases afhangen van vorm informatie en betrouwbaarheid in een stimulus.

Om de resultaten aan bestaande kennis over waarneming te kunnen koppelen, werd gebruik gemaakt van een geometrisch eenvoudige vorm: een oppervlak gevouwen om de horizontale as. Oppervlaktes zijn computer simulaties waarin diepte in een stereogram getoond wordt. Vormwaarneming wordt gemeten met behulp van een *vorm discriminatie taak*. In deze taak worden achtereenvolgens twee gevouwen oppervlaktes getoond en werd de waarnemer gevraagd welke van de twee oppervlaktes de scherpste vouwhoek had. Uit deze vergelijkingen leidden we twee maten af. De eerste maat is het punt waarop de waarnemer twee vormen als gelijk beschouwt, ofwel het punt van subjectieve gelijkenis. Wanneer vorm waarneming veridicaal is, zal dit punt liggen op het moment waarop de vorm van de twee oppervlaktes fysiek gelijk is, maar wanneer waarneming door context beïnvloedt is, zal dit punt afwijken van het punt waarop de twee oppervlaktes fysiek gelijk zijn. De tweede maat is het minimale verschil dat iemand kan waarnemen, ofwel de discriminatie drempel. De discriminatie drempel is een maat voor betrouwbaarheid van vormwaarneming.

Hoofdstuk 2: Waarneming van 3D vorm in context: contrast en assimilatie

Het brein heeft geen toegang tot de driedimensionale structuur van de wereld, en achterhaalt driedimensionale eigenschappen uit projecties op de retina. In deze projecties is diepte informatie niet meer direct aanwezig. Het brein kan diepte reconstrueren uit ‘diepte cues’, zoals binoculaire dispariteit, beweging, perspectief en textuur. Deze cues geven geen volledige informatie over diepte, waardoor de vorminformatie uit individuele cues kan verschillen en geïntegreerd moet worden om tot een bruikbare vormschatting te komen. Vormwaarneming bestaat dus uit verschillende stadia: een waar informatie uit individuele diepte cues verwerkt wordt en een waar informatie uit deze diepte cues geïntegreerd is in een ‘cue invariante’ vormwaarneming. In een eerste experiment onderzochten wij op welk niveau van vormwaarneming contextuele effecten plaatsvinden. Wanneer contextuele effecten plaatsvinden in de verwerking van individuele diepte cues, beïnvloedt de *dispariteitinformatie* in de context de *dispariteitinformatie* in een vorm, maar niet de informatie uit andere diepte cues zoals beweging. Maar wanneer contextuele effecten plaatsvinden op een niveau waar informatie uit de verschillende diepte cues geïntegreerd is, beïnvloeden de twee *vormen* elkaar, ongeacht de cue waarin zij gepresenteerd zijn. Wij maakten van deze tegenstelling gebruik en ontwierpen een stimulus waarin de vorm van een centraal en contextuele oppervlaktes ofwel door dezelfde diepte cue ofwel door een andere diepte cue gedefinieerd was (binoculaire dispariteit en beweging). Wanneer de vorm van de oppervlaktes door de dezelfde diepte cue gedefinieerd was, ontstond een contrast bias. Wanneer hun vorm door een andere diepte cue gedefinieerd was, ontstond een assimilatie bias. Deze resultaten laten zien dat contextuele effecten op meerdere niveaus van vormwaarneming ontstaan. Contrast biases ontstaan op een niveau waar informatie uit individuele diepte cues verwerkt wordt, terwijl assimilatie biases ontstaan op een niveau waar informatie uit verschillende diepte cues geïntegreerd is.

Hoofdstuk 3: Onzekerheid toont omgevingsmodulatie van vorm waarneming

Context informatie kan gebruikt worden om perceptuele onzekerheid over vorm te verminderen. Wanneer je, bijvoorbeeld, de absolute lengte van iemand niet goed kan inschatten, kan je vaak gemakkelijker zien hoeveel langer iemand is dan een aangrenzend persoon. In dit geval heeft informatie uit de context meer invloed naarmate de vorm van de context betrouwbaarder is. In een tweede experiment onderzochten wij hoe een contrast bias afhangt van de betrouwbaarheid van vorm informatie. Betrouwbaarheid van vorm informatie werd beïnvloed door de punten die de oppervlaktes definieerden met een willekeurige vector in de ruimte te verplaatsen zodat de oppervlaktes van een glad vlak in een puntenwolk veranderden. De betrouwbaarheid van ofwel het centrale, de contextuele of beide oppervlaktes werd aangetast waardoor vier betrouwbaarheidscondities ontstonden. Vormwaarneming werd tussen deze condities vergeleken. Wanneer de centrale vorm betrouwbaar was, trad een contrast bias op, die afnam wanneer de vorm van de context onbetrouwbaar was. Wanneer de centrale vorm daarentegen onbetrouwbaar was, trad een assimilatie bias op. Context effecten in vormwaarneming hangen dus af van de betrouwbaarheid van vorm informatie. Het visuele systeem vergroot relatieve verschillen wanneer informatie betrouwbaar is maar integreert wanneer informatie onbetrouwbaar is.

Hoofdstuk 4: Vorm contrast: een globaal mechanisme?

Om de mechanismen van contextuele effecten in vormwaarneming te achterhalen, is het belangrijk te weten of contextuele effecten alleen lokaal of ook globaal plaatsvinden. Wanneer contextuele effecten bijvoorbeeld bepaald worden door globale kenmerken van een stimulus moet op zoek gegaan worden naar een neuraal mechanisme dat toegang tot zulke globale kenmerken heeft. Aangrenzende oppervlaktes beïnvloeden vormwaarneming beïnvloeden, dat is in veel experimenten aangetoond. Maar hoe contextuele informatie over een groter gebied vormwaarneming beïnvloedt, is nog onbekend. In een derde experiment onderzochten wij globale kenmerken van een stimulus een vorm contrast bias beïnvloeden. Daartoe voegden wij aan iedere zijde het centrale oppervlak naast de flankerende oppervlaktes (flankers) een serie van nog vier oppervlaktes toe. De vorm van de flankerende oppervlaktes was constant, maar de vorm verdeling in de serie van aangrenzende oppervlaktes werd gevarieerd. Zo was het lokale vorm verschil met het centrale oppervlak constant maar kon de variatie en het gemiddelde van de globale vormverdeling gevarieerd worden. De gemeten contrast bias bleek niet afhankelijk van het lokale vorm verschil met de flankers maar van het verschil met het gemiddelde van de vorm distributie in de stimulus. De variatie in de vorm distributie, daarentegen, beïnvloedde de contrast bias niet. Deze resultaten laten zien dat vorm contrast veroorzaakt wordt door een mechanisme dat gevoelig is voor globale vorm.

Hoofdstuk 5: Waarneming van 3D helling in perspectief

Twee typen model zijn voorgesteld voor de mechanismen van contextuele effecten in vormwaarneming. Het eerste type model is een statistisch *maximum likelihood estimation* (MLE) model waarin context informatie met directe vorm informatie geïntegreerd wordt om de betrouwbaarheid van vormwaarneming te vergroten. In zo’n MLE model heeft betrouwbare informatie meer invloed op de vormwaarneming en ontstaan contrast biases omdat relatieve informatie de meer betrouwbare informatie is. Het tweede type model is een neuraal model dat beschrijft hoe het brein overtollige informatie verwijdert. In zo’n normalisatie model ontstaan contrast biases doordat de gemiddelde factor uit een stimulus verwijderd wordt, waardoor alleen relatieve verschillen overblijven. In hoofdstuk 5 vergelijken we in hoeverre deze twee modellen in overeenstemming zijn met empirische observaties over contextuele effecten in vormwaarneming. Onze vergelijking toont dat een normalisatie model beter verklaart hoe contextuele effecten afhangen van relatieve vorm verschillen en van de betrouwbaarheid van vorm informatie.

Hoofdstuk 6: Oppervlakte grenzen beperken een diepte na effect niet

Context effecten vinden niet alleen in de ruimte plaats, zoals in de hierboven beschreven experimenten, maar ook in de tijd. In een sequentieel contrast effect (na-effect) leidt langdurige stimulatie met een adaptatie stimulus tot een contrast bias in een opeenvolgend aangeboden stimulus. Deze na-effecten zijn vaak zeer specifiek, wat inhoudt dat zij alleen plaatsvinden tussen twee oppervlaktes die overkomen op een kenmerk (bijvoorbeeld spatiele positie of kleur). De specificiteit van na-effecten kan gebruikt worden om te onderzoeken welke kenmerken in een representatie zijn opgeslagen en hoe visuele informatie georganiseerd is. In een laatste experiment onderzoeken wij in hoeverre een na-effect in diepte-waarneming gerelateerd is aan processen van groepering en segmentatie. Twee typen perceptuele organisatie werden onderzocht. Ten eerste perceptuele organisatie in een coordinaten systeem en ten tweede perceptuele organisatie naar oppervlakte contouren. In een serie experimenten laten wij zien dat een na-effect gevonden kan worden op een oppervlak dat niet overeenkomt met het adaptatie oppervlak in oppervlakte contour of positie. Dit laat zien dat het na-effect veroorzaakt wordt door een minder specifiek mechanisme dan verondersteld.

References

Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review, 101*(3), 414-445.

Anstis, S. M. (1975). What does visual perception tell us about visual coding. In C. Blakemore & M. S. Gazzaniga (Eds.), Handbook of psychobiology (pp. 269-323). New York: Academic Press.

Anzai, A., Ohzawa, I., Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. Nature Neuroscience, 4, 513-518.

Bair, W., Cavanaugh, J. R., & Movshon, A. (2003). Time course and time-distance relationships for surround suppression in macaque v1 neurons *The Journal of Neuroscience, 23*(20), 7690-7701.

Berends, E. M. (2001). Depth perception and adaptation to disparity. Unpublished doctoral dissertation, Utrecht University, Utrecht, The Netherlands.

Berends, E. M., Liu, B., & Schor, C.M. (2005). Stereo-slant adaptation is high level and does not involve disparity coding. *Journal of Vision, 5*(1), 71-80.

Bergman, R. & Gibson, J. J. (1959). The negative aftereffect of the perception of a surface slanted in the third dimension. *American Journal of Psychology, 72*, 364-74.

Blaser, E., & Domini, F. (2002). The conjunction of feature and depth information. *Vision Research, 42*(3).

Bradshaw M. F., Rogers B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. Vision Research, 36, 3457-68.

Brady, T. F. & Chung, M. M. (2007). Spatial constraints on learning in visual search: modelling contextual cuing. Journal of Experimental Psychology: Human Perception and Performance, 33, 798-815.

Brookes, A., & Stevens, K. A. (1989). The analogy between stereo depth and brightness. *Perception, 18*(5), 601-614.

Brown, H. A., Allision, J. D., Samonds, J. M., & Bonds, A. B. (2003). Nonlocal origin of response suppression from stimulation outside the classic receptive field in area 17 of the cat. *Visual Neuroscience, 20*, 85-96.

Carandini, M., Heeger, D. J., & Movshon, A. (1997). Linearity and normalization in simple cells of macaque primary visual cortex. *The journal of neuroscience, 17*(21), 8621-8644.

Cavanaugh, J. R., Bair, W., & Movshon, A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of Neurophysiology, 88*, 2530-2546.

Cornsweet (1970). Visual Perception. Academic Press, New York.

Curran, W. & Johnston, A. (1996). Three dimensional curvature contrast-geometric or brightness illusion? Vision Research, 36, 3641-53.

Domini, F., Adams, W., & Banks, M. (2001). 3D after-effects are due to shape and not disparity adaptation. *Vision Research, 41*.

Domini, F., Blaser, E., & Cicerone, C. M. (2000). Color-specific depth mechanisms revealed by a color-contingent depth aftereffect. *Vision Research, 40*, 359-364.

Domini, F. & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. Journal of Experimental Psychology: Human Perception and Performance, 25, 426-444.

Domini, F., Caudek, C., Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. Vision Research, 46, 1707-1723.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage, 39*, 647-660.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415, 429-433.

Felleman, D. J., van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1, 419-423.

Geisler, W. S., & Albrecht, D. G. (1992). Cortical neurons: Isolation of contrast gain control. *Vision Research, 32*(8).

Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. Journal of Experimental Psychology 16, 1-31.

Gibson. (1979). The ecological approach to visual perception.

Gillam, B., Blackburn, S., & Brooks, K. (2007). Hinge versus twist: The effects of 'reference surfaces' and discontinuities on stereoscopic slant perception. *Perception, 36*, 596-616.

Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of experimental psychology, 14*(2), 163-175.

Gillam, B., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Perception and Psychophysics, 36*(6), 559-564.

Gillam, B., & Pianta, M. J. (2005). The effect of surface placement and surface overlap on stereo slant contrast enhancement. *Vision Research, 45*, 3083-3095.

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

He, Z. J. & Ooi, T. L. (2000). Perceiving depth with reference to a common surface. *Perception, 29*, 1313-1334.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 1992*(9), 181-197.

Graham, M. E. & Rogers, B. J. (1982). Simultaneous and successive contrast effects in the perception of depth from motion parallax and stereoscopic information. Perception 11, 247-262.

Hanada, M. (2004). Effects of the noise level on induced motion. Vision Research, 44, 1757-1763.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth* (Vol. 2): Toronto: I Porteous.

Howe, C.Q., & Purves, D. (2005). Natural-scene geometry predicts the perception of angles and line orientation. PNAS, 102, 1228-1233.

Ishihara, M. (1999). Effect of luminance contrast on the motion aftereffect. *Perceptual Motor Skills, 88*(1), 215-223.

Keck, M. J., Palella, T. D., & Pantle, A. (1976). Motion aftereffect as a function of the contrast of sinusoidal gratings. *Vision Research, 16*(2), 187-191.

Kersten, D. (1987). Predictability and redundancy of natural images. Journal of the optical society of America, 4, 2395-2400.

Knapen, T., & van Ee, R. (2006). Slant perception, and its voluntary control, do not govern the slant aftereffect: Multiple slant signals adapt independently. Vision Research, 46, 3381-3392.

Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research., 43*, 831-854.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgements of surface slant? *Vision Research, 43*, 2539-2558.

Kohn, A. (2007) Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology, 97*, 3155-3164.

Kouh, M., & Poggio, T. (2010). A canonical neural circuit for cortical and nonlinear operations. *Neural Computation, 20*, 1427-1451.

Landy, M. S., Maloney, L.T., Johnston, E. B. , Johnston, M. Y. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Research, 35, 389-412.

Maunsell, J. H., van Essen, D. C. (1983). Functional properties of neurons in the middle temporal visual area of the macaque monkey. II. Binocular interactions and senstivity to binocular disparity. Journal of Neurophysiology, 49, 1148-1167.

Maunsell, J. H., Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. Annual Reviews Neuroscience,10, 363-401.

Mitchell, D. E. & Baker, A. G. (1973). Stereoscopic aftereffects: evidence for disparity-specific neurones in the human visual system.

Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research, 24*(9), 1063-1073.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*(5075), 1357-1363.

Nishina, S., Yazdanbakhsh, A., Watanabe, T., & Kawato, M. (2007). Depth propagation across an illusory surface. *Journal of the optical society of America A., 24*(4), 905.

Noest, A. J., van Ee, R., & van den Berg, A. V. (2006). Direct extraction of curvature-based metric shape from stereo by view modulated receptive fields. *Biological Cybernetics, 95*, 455-486.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature, 5*(391), 580-584.

Poom, L., Olsson, H., Borjesson, E. (2007). Dissociations between slant-contrast and reversed slant-contrast. Vision Research, 47, 746-754.

Ryan, C., & Gillam, B. (1993). A proximity-contingent stereoscopic depth aftereffect: evidence for adaptation to disparity gradients. *Perception, 22*, 403-418.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. M. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience, 2*(8), 733-739.

Schumer, R., & Ganz, L. (1979). Independent stereoscopic channels for different extents of spatial pooling. *Vision Research, 19*(1303-1314).

Schwartz, O., Hsu, A. & Dayan, P. (2007). Space and time in visual context. Nature Neuroscience Reviews, 8, 522-535.

Schwartz, O., Sejnowski, T., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision, 9*, 1-20.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience, 4*(8), 819-826.

Sengpiel, F., Sen, A., & Blakemore, C. (1997). Characteristics of surround inhibition in cat area 17. *Experimental Brain Research, 116*, 216-228.

Series, P., Lorenceau, J., & Fregnac, Y. (2003). The "silent" surround of V1 receptive fields: theory and experiments. *Journal of Physiology - Paris, 97*, 453-474.

Shipley, T., & Hyson, M. (1972). The stereoscopic sense of order--a classification of stereograms. *American Journal of Optometry & Archives of American Academy of Optometry, 49*(2), 83-96.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28*(3), 371-386.

Tajima, S., Watanabe, T., Imai, C., Ueno, K., Asamizuya, T., Sun, P., et al. (2010). Opposing effects of contextual surround in human early visual cortex revealed by functional magnetic resonance imaging with continuously modulated visual stimuli. *The journal of neuroscience, 30*(9), 3264-3270.

Taya, S., Sato, M., & Nakamizo, S. (2005). Stereoscopic depth aftereffects without retinal position correspondence between adaptation and test stimuli. *Vision Research, 45*, 1857-1866.

Thaler, L., & Goodale, M. A. (2010). Beyond distance and direction: The brain represents target locations non-metrically. *Journal of Vision, 10*(3), 1-27.

Te Pas, S. F., & Kappers, A. M. L. (2001). First-order structure induces the 3-D curvature contrast effect. Vision Research, 41, 3829-3835.

Te Pas, S. F., Rogers, B. J. & Ledgeway, T. (2000). A curvature contrast effect for stereoscopically defined surfaces. Current Psychology Letters: Brain, Behaviour and Cognition, 1, 117-126.

Tjan, B. S., Lestou, V., & Kourtzi, Z. (2006). Uncertainty and invariance in the human visual cortex. Journal of neurophysiology, 96, 1556-1568.

Todd, J. T, Thaler, L. & Dijkstra, T. M. H. (2005). The effects of field of view on the perception of 3D slant from texture. Vision Research, 45, 1501-1517.

Treder, M. S. & Meulenbroek, R. G. J. (2010) Integration of structure-from-motion and symmetry during surface perception. *Journal of Vision, 10*(4), 1-17.

Tsutsui, K. I., Sakata, M. T., Naganuma, T., Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298*(11), 409-377.

van der Kooij, K., & Te Pas, S. F. (2009a). Perception of 3D shape in context: Contrast and assimilation. *Vision Research, 49*, 746-751.

van der Kooij, K., & Te Pas, S. F. (2009b). Uncertainty reveals surround modulation of shape. *Journal of Vision, 9*(3), 1-8.

van Ee, R., Banks, M. S., Backus, B. T. (1999). An analysis of binocular slant contrast. Perception, 28, 1121-1145.

van Ee, R., & Erkelens, C. J. (1996). Anisotropy in Werner’s binocular depth-contrast effect. *Vision Research, 36*(35), 2253-2262.

Van Essen, D.C., Newsome, W.T., Maunsell, J.H.R. (1984). The visual-field representation in striate cortex of the macaque monkey –asymmetries, anisotropies, and individual variability. Vision Research, 24, 429-448.

van Lier, R., & Wagemans, J. (1997). Perceptual grouping measured by color assimilation: Regularity versus proximity. Acta Psychologica 97, 37-70.

Welchmann, A. E., Deubelius, A., Conrad, V., Bulthoff, H. & Kourtzi, Z. (2005). 3d shape perception from combined depth cues in human visual cortex. Nature Neuroscience, 8, 820-827.

Westheimer, G. (1990). Simultaneous orientation contrast for lines in the human fovea. *Vision Research, 30*(11), 1913-1921.

Wilson, H. R., & Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision Research, 33*(8), 1133-1149.

Zeki, S.M. (1978). Uniformity and diversity of structure and function in rhesus-monkey prestriate visual-cortex. Journal of Physiology, 277, 273-290.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience, 16*(22), p. 7376-7389.

Acknowledgments

In the past chapters of this thesis, many articles have been cited that provided the scientific ground on which my experiments and conclusions have been built. In addition, many informal discussions and experiences have allowed the research in this thesis to grow.

All experiences during my PhD project have been set in the context of my daily interaction with Susan te Pas, my advisor. Her research proposal turned out to be a fruitful ground for my experiments, which immediately taught me that all experimental work starts with a well-thought idea. She created a supportive environment where I could learn freely, at the same time always giving me critical feedback. Patiently, she showed me the power of a more exact approach.

Frans Verstraten, my professor, taught me to always keep in mind the significance of the subject I am investigating. He also showed me the importance of forming friendships in the scientific community and learning from them. The first experience of such a friendship was with Tomas Knapen and Jan Brascamp, with whom I founded the newsletter of the Helmholtz Research institute, the Helmholtz Herald. Discussing scientific ideas and problems in a social setting freed my mind and opened me up to new ideas. Jan and Ans Koenderink, the founders of my research school, seemed wizards of psychophysics when I started my PhD. I never new that wizards would be so generous in sharing their knowledge.

At the same time I have to thank someone who took an entirely different approach in helping me grow. Ignace Hooge frustrated me up to breaking point in weekly lab meetings. Now, I have to admit that battling the continuous interruptions have strengthened my skills in having a scientific discussion that rapidly comes to the point. However, this approach wouldn’t have been successful without the support of Mijke and Titia, my paranymphs. They are both strong ladies and inspired me to take on any challenge.

A few papers stood out in how they inspired me, and I have been lucky enough to visit the authors of these papers. Fulvio Domini provided my first experience with a lab visit. The week I spend at Brown University fuelled my enthusiasm to move beyond the phenomenology of vision. In search of the mechanisms of perception, I came across a paper by Odelia Schwartz. Odelia welcomed me at Einstein College and helped me to develop a more thorough understanding of her model of contextual biases in visual perception. Also many thanks to Johan Wagemans, Pascal Mamassian and Mark Wexler, who welcomed me in their lab and asked me protruding questions which helped me to find the important issues in science.

In these acknowledgments my family and friends cannot be forgotten. They are all curious, original and strong-minded characters that have inspired me to always follow my own path and make the most of the opportunities that a PhD project offers.

Publications

Journal Papers

van der Kooij, K., Domini, F. & Te Pas, S. F. Surface Boundaries Do Not Constrain A Depth Aftereffect. Under revision at Vision Research.

van der Kooij, K. & te Pas, S. F. Shape Contrast: A Global Mechanism? Accepted for publication in Vision Research.

van der Kooij, K. & te Pas, S. F. (2009) Perception of 3D Shape in Context: Contrast and Assimilation. Vision Research, 49, p. 746-751.

van der Kooij, K. & te Pas, S. F. (2009) Uncertainty reveals surround modulation of shape. Journal of Vision, 3, 9, p. 1-8.

William L. Thompson, Stephen M. Kosslyn, Michael S. Hoffman & Katinka van der Kooij (2008): Inspecting Visual Mental Images: Can People “See” Implicit Properties as Easily in Imagery and Perception? Memory and Cognition.

Refereed Abstracts

van der Kooij, K. & te Pas, S. F. (2010) Shape Contrast: Where Does It End? Journal of Vision 10(7), 51.

van der Kooij, K. & te Pas, S. F. (2009) Surface boundaries do not constrain a depth aftereffect. Perception 38 ECVP Abstract Supplement, 184.

van der Kooij, K. & te Pas, S. F. (2009) Figure-plane binds the depth aftereffect. Journal of Vision 9(8), 66.

van der Kooij, K. & te Pas, S. F. (2008) Shape Surround Interactions Cross Cue Modalities. Perception 37 ECVP Abstract Supplement, 124.

van der Kooij, K., & te Pas, S. F. (2008). Contextual bias of slant perception in unreliable context. Journal of Vision, 8(6), 844

van der Kooij, K. & te Pas, S. F. (2007) Curvature contrast governed by spatial proximity, not depth cue. Perception 36 ECVP Abstract Supplement.

van der Kooij, K. & te Pas, S. F. (2007) Curvature contrast occurs after cue combination. Journal of Vision 7(9), 1534-7362.

van der Kooij, K. & te Pas, S. F. (2007) Contextual interactions constrained by spatial proximity and depth cue. NVP Winterconference Abstract Supplement.

Curriculum Vitae

Katinka van der Kooij was born in Wageningen (the Netherlands) on the 4th of August 1981 but went to highschool in the Hague (Christelijk Gymnasium Sorghvliet). After highschool she studied for one year as a liberal arts undergraduate at Wittenbergh University, Springfield Ohio. She subsequently studied Psychology at the University of Amsterdam, carrying out a 4-month internship at the Kosslyn laboratory at Harvard University. She graduated in 2004 and kept working as a research assistant at the University of Amsterdam until she started her PhD position at Utrecht University in May 2006. There she worked under the supervision of professor Frans Verstraten and dr. Susan te Pas. As part of her PhD program she made research visits to professor Fulvio Domini's lab at Brown University in Providence, Rhode Island and to professor Odelia Schwartz at Einstein College in New York. Parallel to her scientific work she has been involved in creative writing, which resulted in the publication of a collection of short stories in October 2008 at publishing house Prometheus.