

# An in-depth vision on perception neural mechanisms of shading and motion object cue processing

Citation for published version (APA):

Gräwe, B. (2013). *An in-depth vision on perception neural mechanisms of shading and motion object cue processing*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20130618bg>

## Document status and date:

Published: 01/01/2013

## DOI:

[10.26481/dis.20130618bg](https://doi.org/10.26481/dis.20130618bg)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# **An In-Depth Vision on Perception**

## **Neural Mechanisms of Shading and Motion Object Cue Processing**

Britta Graewe

## Colophon

Layout&Cover Design: Britta Graewe  
Production: Datawyse, Universitaire Pers Maastricht  
Dutch translation: Marin Been, Bert Jans  
Copyright: © Britta Graewe, Maastricht 2013  
ISBN number: 978-90-9027624-3

# **An In-Depth Vision on Perception**

## **Neural Mechanisms of Shading and Motion Object Cue Processing**

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof dr L.L.G. Soete  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op dinsdag 18 Juni 2013 om 14.00 uur

door

**Britta Graewe**

**Promotores**

Prof. dr. P. De Weerd

Prof. dr. M. Castelo-Branco (Universidade de Coimbra, Portugal)

**Beoordelingscommissie**

Prof. dr. B. Jansma (voorzitter)

Prof. dr. E. Formisano

Prof. dr. R. Goebel

Dr. L. M. Jonkman

Dr. A. L. Kaas

## CONTENTS

---

	<b>PROLOGUE</b>	7
<b>Chapter 1</b>	3D Shading and Motion Cue Combinations: Suppressive Competition instead of Integration	19
<b>Chapter 2</b>	An Exploration of the Neural Representations of Combined Shading and Motion Cues in Human Visual Cortex	47
<b>Chapter 3</b>	Stimulus Dependency of Object-Evoked Responses in Human Visual Cortex: An Inverse Problem for Category Specificity	79
<b>Chapter 4</b>	Impaired Processing of 3D Motion-Defined Faces in MCI and Healthy Ageing: An fMRI Study	103
	<b>EPILOGUE</b>	131
	Summary	137
	Savenvatting	139
	Acknowledgements	143
	Curriculum Vitae	147



## PROLOGUE

Understanding the link between brain and perception is one of the most exciting neuro-scientific endeavours. Our perceptual system provides us with impressions about the external world by creating a representation of the environment in our minds. But this process is far from trivial. Consider the visual modality: Perception starts with transient light patterns that fall on receptors in the retina, creating an inverted 2D image of our surroundings. By a cascade of electrical signals emanating from photoreceptors, this information travels via interconnected neuronal pathways from the eye to several areas of the brain where eventually a stable and coherent 3D percept is re-constructed.

The segregation of foreground from background, and more generally, the recovery of depth from an essentially flat projection of the world on the retina is one of the challenges the visual system faces. We are remarkably apt in solving this problem however, and to that aim exploit both monocular depth cues (shading, texture and motion) as well as binocular depth cues (disparity). The visual system's tricks have fascinated artists and scientists alike. While a substantial amount of our knowledge of the underlying mechanisms comes from early psychophysical studies, new techniques such as the invention of functional magnetic resonance imaging (fMRI), electroencephalography (EEG) as well as electrophysiological recordings in primates have enabled scientists to study the exact neural locus and timing underlying 3D visual perception. The present work will focus on the neural foundations of 3D form perception from shading and motion visual cues. Three research topics will be addressed in 4 studies.

1. The mechanisms underlying the combination of multiple visual cues for the perception of simple 3D forms (**Chapters 1 and 2**)
2. The relationship between cue-invariant high-level object vision and ERP components (**Chapter 3**)
3. 3D structure-from-motion (SFM)-defined face perception as a tool to dissociate healthy from pathological visual processing (**Chapter 4**)

The present Prologue provides a brief *General background* for each of these three research topics, after which a *Thesis overview* is given in which the studies that were performed will be introduced.



*Dalmatian, Photograph by R.C. James*

## **General Background**

Consider the illustration of a Dalmatian above. Based on the viewer's world knowledge (experience with leaves, dogs, shadows etc), the black and white shapes in the image can become perceptually organized. This perceptual process is embedded in the structure and function of the brain, shaped by genetic and experience-driven constraints. While the use of texture and shading cues may suffice to distinguish the figure from its background, in real life the dog's motion would provide an additional strongly disambiguating cue with regard to its location and 3D shape. But what are the neural mechanisms that eventually enable us to perceive a Dalmatian playing in the park from moving 3D black and white image fragments whose projection hits the back of our eye?

From the retina information about objects is relayed to the lateral geniculate nucleus (LGN) of the thalamus and subsequently travels to the primary visual cortex in the occipital lobe. Starting from here, two major high level visual cortical pathways mediate perception: A ventral pathway concerned with object identification ('what?') projecting to the inferior temporal cortex and a dorsal pathway processing spatial and motion stimulus properties ('where') projecting to parietal areas (Ungerleider and Mishkin 1982; Martin et al. 1995; Ishai et al. 2000). The shape of the Dalmatian in our example is described by shading cues, a predominantly ventral stream process (Georgieva et al. 2008), as well as structure-from-motion (SFM) cues which are predominantly computed in the dorsal stream with comparatively less contributions of ventral stream regions (Orban et al. 2006). In addition, texture depth cues, which require both visual streams (Georgieva et al. 2008; Nelissen et al. 2009) can be exploited. However, the question arises whether

the coherent perception of shape from several different visual cues would not necessitate interaction between different functional streams and anatomical pathways. Effectively, while the concept of - to a certain extent - independent, parallel visual streams is an elegant view and accommodates many psychophysical, neurophysiological, neuroanatomical, neuropsychological as well as neuroimaging findings, many challenges have been raised. In fact, a large number of publications have suggested intricate interactions between ventral and dorsal processing pathways, both in lower- and higher-order visual areas (Merigan and Maunsell 1993; Sincich and Horton 2005; Farivar 2009; Kravitz et al. 2012).

### ***Topic 1: Processing of multiple visual image cues for the perception of simple 3D forms***

The question of integration between different functional streams also pertains to the specific cues of interest in the present thesis – namely the different monocular cues from which we derive depth (3D). Current knowledge suggests that information from different depth cues becomes integrated, generally resulting in improved perceptual sensitivity compared to single cues (Ernst and Bühlhoff 2004; Lovell and Bloj 2012). Combinations of visual depth cues such as shading and disparity (Zhang et al. 2007), motion and disparity (Bradshaw and Rogers 1996), texture and disparity (Knill and Saunders 2003) as well as texture and motion (Young et al. 1993) have been studied extensively. Most depth cues within- and across modalities combine cooperatively according to a maximum-likelihood estimation (MLE) principle, with combined cue estimates being more reliable than single cue estimates. Moreover, according to MLE, the weighting of each cue in the composite percept is based on single cue reliabilities (Ernst and Banks 2002).

This beneficial behavioural effect of combining cues implies that after initial processing in relatively segregated cue specific regions (Zeki et al. 1991; Tootell et al. 1995; Sincich and Horton 2005), integration at a neural level takes place. Neurophysiological recording studies in primates have shown evidence for inferior temporal cortex neurons with stimulus property representations that generalize over texture and disparity depth cues (Liu et al. 2004), disparity and luminance cues (Tanaka et al. 2001) as well as over 2D luminance and relative motion cues (Vogels and Orban 1996). Moreover, the surprisingly few human neuroimaging studies that have confronted this topic to date, have reported cue integration for combined disparity and motion cues in visual area V3B (Ban et al. 2012) as well as combined disparity and perspective cue representations in extrastriate areas LO and MT (Welchman et al. 2005). In addition, generic, cue-invariant representations of depth have been reported in the IPS (depth-from shading: Taira et al. 2001; depth-from-motion: Kriegeskorte et al. 2003), which makes this region also a particularly suitable candidate for cue combinations to take place.

### ***Topic 2: Cue invariant, high-level object representations***

Neurons in increasingly higher-order areas related to object processing are likely to permit cue integration for an increasing number of cues and for object features of

increasing complexity (Welchman et al. 2005). In addition to the statistical property of enhanced perception due to cue integration, cue integration also relates to cue invariance. Cue invariance refers to neural responses that are fairly robust with respect to the visual cue that defines them. Hence, visual areas that show cue invariance can be thought of as good candidates for optimal cue integration to take place.

In monkeys, a variety of studies have shown cue invariance in high-level areas for complex visual objects (Sary et al. 1993; Sereno et al. 2002). In humans, ventral stream shape and object representations such as specialized regions for faces (Kanwisher and Yovel 2006), houses (Epstein and Kanwisher 1998) or words (Cohen et al. 2000) have been shown by numerous fMRI studies. Many of these representations have indeed been shown to be invariant with respect to the 3D depth cue that describes them (Grill-Spector et al. 1998; Kriegeskorte et al. 2003). Category-specificity for stimuli defined by different cues has also been investigated using event-related potentials (ERPs), with the most famous examples being the face specific potentials P100 (Herrmann et al. 2005) and N170 (Bentin et al. 1996; Itier 2004). Compared to fMRI, ERP recordings bear more precise temporal information on stimulus processing at the expense of spatial specificity. Importantly, although the face specific N170 has been found in response to various face and object representations (Rossion and Jacques 2011), its amplitude was also manipulated by factors unrelated to a specific category (e.g. spatial frequency (Goffaux et al. 2003), inter-stimulus perceptual variance (Thierry et al. 2007b), stimulus orientation (Jacques and Rossion 2007) and cue type (Latinus and Taylor 2006)). Based on these findings, the face-specificity of the N170, and more generally the idea of specific categorical information in ERP amplitudes has been a matter of strong debate (Bentin et al. 2007; Thierry et al. 2007a, 2007b).

### ***Topic 3: 3D SFM-defined faces: a sensitive tool to dissociate normal from pathological functioning***

While precise understanding of normal perceptual functioning is important, it also enables us to understand perceptual losses due to ageing, disease or brain damage. Alzheimer's disease (AD) is one of the most common and serious medical conditions of later life. In addition to memory complaints, AD is characterized by several deficits in high level visual functions. While impairments include ventral stream dysfunctions such as face (Sauer et al. 2006) and object recognition (Cronin-Golomb et al. 1995), the most predominant deficits have been reported in the visuospatial (Mendez et al. 1996) and motion perception domain (Mapstone et al. 2006; Kavcic et al. 2011). Several structural imaging studies in AD patients have extended and corroborated these findings of behavioural deficits, showing reductions in grey matter in several mediotemporal, parietal as well as frontal regions (Baron et al. 2001; Pennanen et al. 2005; Singh et al. 2006). Overall, the majority of findings described in the literature point to the visual dorsal stream as a major target of neuropathologies in the course of the disease. Importantly, specific dorsal stream dysfunctions have already been demonstrated in mild cognitive impairment

(MCI), a boundary cognitive stage between normal ageing and AD (Petersen et al. 1999; Mapstone et al. 2003). In MCI, among the several subtypes that have been identified, in particular amnesic MCI patients show an accelerated conversion to AD (yearly transition rate of 10–15% (Petersen et al. 2001)). Despite the lack of a prevention or cure for AD, early diagnosis followed by pharmacological or cognitive therapies may succeed to halt or slow down progression of the disease. Therefore, detecting small perceptual deficits related to early stages of neuropathology may be an important goal which can facilitate early diagnosis and allow for timely treatment of true preclinical AD. In addition, visual perception tasks may serve as valuable biomarkers of pharmacological treatment effects (Bokde et al. 2009).

Previously, a high correlation between SFM perception deficits and MCI (Lemos et al. 2011) as well as AD severity (Kim and Park 2010) has been shown. Since in SFM, motion perception in the dorsal stream provides the basis for object perception in the ventral stream, this suggests that not merely dorsal stream functions but also dorsal-ventral stream interactions may be affected in MCI patients. Thus, meaningful SFM object stimuli may provide a particularly sensitive tool for investigating perceptual deficits in MCI by simultaneously assessing both dorsal stream functions and ventral stream-related object perception.

## **Thesis overview**

We have used a multi-disciplinary approach using psychophysics, EEG and fMRI to study the mechanisms underlying 3D vision from shading and motion visual cues. In chapter 1 we behaviourally investigated depth perception from single shading and single motion cues as well as from consistent and inconsistent cue composites. In chapter 2 we continued this investigation using fMRI to gain insight in the neural locus of combined shading-motion perception. Findings are placed in the context of computational models and frameworks for cue combination.

Object representations in inferior temporal cortex have been shown both using fMRI and ERP recordings. In chapter 3 we assessed the use of EEG as a method for investigating *cue-invariant* 3D face representations defined by luminance and motion cues. The work described in chapter 4 addresses fMRI signal alterations underlying deficits in 3D structure-from-motion face perception in the healthy ageing and the MCI brain. Distinctive face-related visual activation patterns that allow differentiation between ageing and MCI participants are demonstrated.

## **Chapter 1 & 2 (Topic 1)**

In the first two chapters we addressed the mechanisms underlying shading and motion cue combinations and the circumstances that are required for cues to combine in a cooperative fashion (Topic 1).

The combination of multiple consistent depth cues has previously been shown to improve surface perception, leading to more reliable estimates compared to single cues (Hillis et al. 2004). Here, the motivation to study the effects of combining shading and motion visual depth cues was driven by the virtually complete absence of information on the effect of 3D shading-motion cue combinations in the

literature. We therefore examined whether cue integration according to MLE (Clark and Yuille 1990) could also be demonstrated for this specific cue pair. To that end, we employed a curvature discrimination task based on single shading- and single motion-defined shapes as well as based on combined cue shapes in which curvature was defined by both cues in either a consistent or inconsistent manner. This design enabled us to compare curvature sensitivity across the different cue conditions. While chapter 1 focused on studying behavioural effects of shading-motion combination, in chapter 2, we aimed to extend this investigation using event-related fMRI coupled with a curvature discrimination task. Thus, in chapter 2 we sought to gather information on the neural underpinnings of the observed behavioural effects in shading-motion combined conditions. However, the analysis described in this chapter is still on-going, which constrains the conclusions that can be drawn at this point in time.

### **Chapter 3**

In chapter 3 we examined the relationship between cue-invariant high-level object vision and ERP components. A specialization for faces in the brain has been studied extensively using the “face-specific” electrophysiological component N170 (Bentin et al. 1996). However, the characteristics of the N170 in response to structure-from-motion (SFM) face stimuli have, to our knowledge, not yet been investigated. These stimuli are interesting for the purpose of the present study since the only cue to face perception is provided by a moving dot pattern (Andersen and Bradley 1998), containing fundamentally different low level feature properties than commonly used static face stimuli. Here, the use of SFM and luminance-defined face and chair stimuli enabled us to test two research questions: First, are the face-specific effects seen at the N170 cue invariant and can they also be manifested when faces are defined by motion? Secondly, do manipulations of the stimulus depth level while leaving the object category percept intact affect the amplitude of the N170? If the N170 amplitude was not only modulated by the stimulus category but was also susceptible to these visual stimulus feature manipulations, the implication would be that the N170 does not reliably carry category-specific information.

### **Chapter 4**

Here, we investigated the use of 3D SFM-defined face perception as a tool to dissociate healthy from pathological visual processing. Prior psychophysical studies have demonstrated a high correlation between SFM perception deficits and MCI (Lemos et al. 2011) as well as AD severity (Kim and Park 2010). In the presented study, we used concurrent fMRI and an SFM face categorization task to understand the neural basis of these behavioural impairments in the healthy aging and MCI. SFM stimuli, with motion integration mechanisms in the dorsal stream forming the basis for object recognition in the ventral stream, provide a sensitive tool to probe dorsal–ventral cross-talk. Moreover, the use of SFM cues to describe face stimuli, provided an additional window on temporal lobe processes related to face percep-

tion (Kanwisher et al. 1997; Kriegeskorte et al. 2003). By manipulating stimulus depth and duration levels we aimed to increase the sensitivity of our paradigm to differentiate both psychophysical performance and associated fMRI brain activation between MCI-afflicted and normal aging participants.

## **Research questions in this thesis**

Based on the background and overview of questions offered in the preceding paragraphs, the 3 main research questions addressed in 4 separate studies can be formulated as follows:

1. What are the mechanisms underlying the combination of shading and motion visual cues to 3D perception and where in the brain does this combination take place? (chapter 1 & chapter 2)
2. Can we separate categorical perception from stimulus-specific effects based on the amplitude of object-evoked ERP components? (chapter 3)
3. How are the neural underpinnings of 3D vision and categorical perception affected by healthy ageing and mild cognitive impairment (MCI)? (chapter 4)

## References

- Andersen RA, Bradley DC. 1998. Perception of three-dimensional structure from motion. *Trends Cogn Sci.* 2:222–228.
- Ban H, Preston TJ, Meeson A, Welchman AE. 2012. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci.* 15:636–643.
- Baron J-C, Chételat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F. 2001. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage.* 14:298–309.
- Bentin S, Allison T, Puce A, Perez E, McCarthy G. 1996. Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci.* 8:551–565.
- Bentin S, Taylor M, GA R, Itier R, Caldara R, Schyns P, Jacques C, Rossion B. 2007. Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nat Neurosci.* 10:801–802.
- Bokde ALW, Karmann M, Teipel SJ, Born C, Lieb M, Reiser MF, Möller H-J, Hampel H. 2009. Decreased activation along the dorsal visual pathway after a 3-month treatment with galantamine in mild Alzheimer disease: a functional magnetic resonance imaging study. *J Clin Psychopharm.* 29:147–156.
- Bradshaw MF, Rogers BJ. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res.* 36:3457–3468.
- Clark J, Yuille A. 1990. *Data fusion for sensory information processing systems.* Kluwer, Academic Publishers.
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff M a, Michel F. 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain.* 123 ( Pt 2:291–307.
- Cronin-Golomb A, Corkin S, Growdon JH. 1995. Visual Dysfunction predicts cognitive deficits in Alzheimer's disease. *Optometry Vision Sci.* 72:168–176.
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature.* 392:598–601.
- Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature.* 415:429–433.
- Ernst MO, Bühlhoff HH. 2004. Merging the senses into a robust percept. *Trends Cogn Sci.* 8:162–169.
- Farivar R. 2009. Dorsal-ventral integration in object recognition. *Brain Res Rev.* 61:144–153.
- Georgieva SS, Todd JT, Peeters R, Orban G a. 2008. The extraction of 3D shape from texture and shading in the human brain. *Cereb Cortex.* 18:2416–2438.

- Goffaux V, Gauthier I, Rossion B. 2003. Spatial scale contribution to early visual differences between face and object processing. *Brain Res Cogn Brain Res*. 16:416–424.
- Grill-Spector K, Kushnir T, Edelman S, Itzhak Y, Malach R. 1998. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*. 21:191–202.
- Herrmann MJ, Ehlis A-C, Ellgring H, Fallgatter a J. 2005. Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *J Neural Transm*. 112:1073–1081.
- Hillis J, Watt S, Landy M, Banks M. 2004. Slant from texture and disparity cues: Optimal cue combination. *J Vis*. 4:967–992.
- Ishai A, Ungerleider LG, Martin a, Haxby JV. 2000. The representation of objects in the human occipital and temporal cortex. *J Cogn Neurosci*. 2:35–51.
- Itier RJ. 2004. N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs. *Cereb Cortex*. 14:132–142.
- Jacques C, Rossion B. 2007. Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *Neuroimage*. 36:863–876.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.
- Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos T Roy Soc B*. 361:2109–2128.
- Kavcic V, Vaughn W, Duffy CJ. 2011. Distinct visual motion processing impairments in aging and Alzheimer's disease. *Vision Res*. 51:386–395.
- Kim N-G, Park J-H. 2010. Learning to perceive structure from motion and neural plasticity in patients with Alzheimer's disease. *Neuropsychologia*. 48:1464–1471.
- Knill DC, Saunders J a. 2003. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res*. 43:2539–2558.
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2012. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci*. 17:26–49.
- Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. *J Neurosci*. 23:1451–1463.
- Latinus M, Taylor MJ. 2006. Face processing stages: impact of difficulty and the separation of effects. *Brain Res*. 1123:179–187.
- Lemos R, Figueiredo P, Santana I, Simões MR, Castelo-Branco M. 2011. Temporal Integration of 3D Coherent Motion Cues Defining Visual Objects of Unknown Orientation is Impaired in Amnesic Mild Cognitive Impairment and Alzheimer's Disease. *J Alzheimers Dis*. 27:1–12.

- Liu Y, Vogels R, Orban G a. 2004. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *J Neurosci.* 24:3795–3800.
- Lovell P, Bloj M. 2012. Optimal integration of shading and binocular disparity for depth perception. *J Vis.* 12:1–18.
- Mapstone M, Logan D, Duffy CJ. 2006. Cue integration for the perception and control of self-movement in ageing and Alzheimer's disease. *Brain.* 129:2931–2944.
- Mapstone M, Steffenella TM, Duffy CJ. 2003. A visuospatial variant of mild cognitive impairment: getting lost between aging and AD. *Neurology.* 60:802–808.
- Martin A, Haxby J, Lalonde F. 1995. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science.* 270:102–105.
- Mendez M, Cherrier M, Meadows R. 1996. Depth perception in Alzheimer's disease. *Percept Mot Skills.* 83:987–995.
- Merigan WH, Maunsell JH. 1993. How parallel are the primate visual pathways? *Annu Rev Neurosci.* 16:369–402.
- Nelissen K, Joly O, Durand J-B, Todd JT, Vanduffel W, Orban G a. 2009. The extraction of depth structure from shading and texture in the macaque brain. *PloS one.* 4:e8306.
- Orban G a, Claeys K, Nelissen K, Smans R, Sunaert S, Todd JT, Wardak C, Durand J-B, Vanduffel W. 2006. Mapping the parietal cortex of human and non-human primates. *Neuropsychologia.* 44:2647–2667.
- Pennanen C, Testa C, Laakso MP, Hallikainen M, Helkala E-L, Hänninen T, Kivipelto M, Könönen M, Nissinen A, Tervo S, Vanhanen M, Vanninen R, Frisoni GB, Soininen H. 2005. A voxel based morphometry study on mild cognitive impairment. *J Neurol Neurosur Ps.* 76:11–14.
- Petersen R, Smith G, Waring S, Ivnik R, Tangalos E, Kokmen E. 1999. Mild cognitive impairment. *Arch Neurol.* 56:303–309.
- Petersen RC, Doody R, Kurz A, Mohs R, Morris J, Rabins P, Ritchie K, Rossor M, Thal L, Winblad B. 2001. Current concepts in mild cognitive impairment. *Arch Neurol.* 58:1985.
- Rossion B, Jacques C. 2011. The N170: understanding the time-course of face perception in the human brain. In: Luck S, Kappenman E, editors. *The Oxford handbook of ERP componenets.* Oxford, Uk: Oxford University press.
- Sary G, Vogels R, Orban G. 1993. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science.* 260:995–997.
- Sauer J, Ffytche DH, Ballard C, Brown RG, Howard R. 2006. Differences between Alzheimer's disease and dementia with Lewy bodies: an fMRI study of task-related brain activity. *Brain.* 129:1780–1788.
- Sereno ME, Trinath T, Augath M, Logothetis NK. 2002. Three-Dimensional Shape Representation in Monkey Cortex. *Neuron.* 33:635–652.

- Sincich LC, Horton JC. 2005. The circuitry of V1 and V2: integration of color, form, and motion. *Annu Rev Neurosci.* 28:303–326.
- Singh V, Chertkow H, Lerch JP, Evans AC, Dorr AE, Kabani NJ. 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer’s disease. *Brain.* 129:2885–2893.
- Taira M, Nose I, Inoue K, Tsutsui K. 2001. Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. *Neuroimage.* 14:959–966.
- Tanaka H, Uka T, Yoshiyama K, Kato M, Fujita I. 2001. Processing of Shape Defined by Disparity in Monkey Inferior Temporal Cortex. *JN Physiol.* 85:735–744.
- Thierry G, Martin C, Downing P, Pegna A. 2007a. Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nat Neurosci.* 10:802–803.
- Thierry G, Martin CD, Downing P, Pegna AJ. 2007b. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci.* 10:505–511.
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci.* 15:3215–3230.
- Ungerleider LG, Mishkin M. 1982. Two cortical visual systems. In: Ingle D, Goodale R, R JW M, editors. *Analysis of visual behavior.* Cambridge, MA: MIT. p. 549–586.
- Vogels R, Orban G a. 1996. Coding of stimulus invariances by inferior temporal neurons. *Progr Brain Res.* 112:195–211.
- Welchman AE, Deubelius A, Conrad V, Bülhoff HH, Kourtzi Z. 2005. 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci.* 8:820–827.
- Young MJ, Landy MS, Maloney LT. 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.* 33:2685–2696.
- Zeki S, Watson J, Lueck C. 1991. A direct demonstration of functional specialization in human visual cortex. *J Neurosci.* 17:641–649.
- Zhang Y, Weiner VS, Slocum WM, Schiller PH. 2007. Depth from shading and disparity in humans and monkeys. *Visual Neurosci.* 24:207–215.



## 3D Shading and Motion Cue Combinations: Suppressive Competition instead of Integration

---

*Based on: Graewe B, Bastiani M, Duecker F, Roebroek A, Castelo-Branco M, De Weerd P. 3D shading and motion cue combinations: Suppressive competition instead of integration (in preparation).*

## Abstract

The integration of multiple visual depth cues has been shown to improve three-dimensional (3D) shape estimation according to a maximum-likelihood estimation (MLE) model. MLE predicts higher reliability for combined compared to single cue estimates, with a bias of combined-cue estimates towards the most reliable single cue estimate. Hence, relative cue reliability sets cue weights during their combination. Here, we tested this model for the virtually unstudied combination of visual 3D shading and motion cues. Human participants discriminated 3D curvatures of ellipsoidal test shapes from reference shapes, defined by shading, motion or combined cues. While curvature in motion-defined test shapes was *underestimated*, and curvature in shading-defined test shapes was *overestimated*, consistent combined cue estimates were unbiased. Surprisingly, curvature estimates were *less* reliable for combined than single cues, suggesting a form of competition which suppresses the reliability of combined cue judgments. Moreover, psychometric functions and response times in the combined condition represented a weighted average of judgments based on either cue alone, with a greater weight assigned to the shading cue. The weak influence of the motion cue was also observed when shading and motion cues defined inconsistent curvatures with more accurate fixation decreasing the weight of the motion cue even further. Importantly, cue weighting could not be accounted for by the individual cues' reliabilities. The lack of shading-motion integration supports the idea that cue integration is conditional on constraints, implemented in the visual system's architecture, and reflecting prior world knowledge in line with a Bayesian theoretical framework.

## Introduction

The human visual system reconstructs 3D shapes from 2D retinal cues by exploiting a wide range of visual depth cues such as texture, shading, motion, and stereopsis. While each of these cues has been studied in isolation (Orban, 2011 for a review), an interesting question pertains to how and under what conditions they are integrated. Previous studies have shown that objects defined by combined consistent cues typically produce perceptual judgments that are more accurate (i.e., less biased), more reliable (i.e., less variable), and more rapid (i.e. shorter response times) than objects defined by single cues. Benefits from combining cues have been confirmed for visual cue combinations of shading and disparity (Vuong et al. 2006; Zhang et al. 2007; Lee and Saunders 2011), motion and disparity (Johnston et al. 1994; Bradshaw and Rogers 1996), texture and disparity (Knill and Saunders 2003), texture and motion (Young et al. 1993), two- and threefold combinations of disparity, shading and motion parallax (Schiller et al. 2011) as well as for multisensory cue combinations (Hillis et al. 2002).

A common statistical framework for this form of “cue cooperativity” (Bülthoff and Mallot, 1987), assumes a two-stage process during which estimates of individual cues are derived separately before a weighted combination that produces a reliability superior to that of the contributing individual cues is calculated (maximum likelihood estimation (MLE), Clark and Yuille (1990)). In addition, the weights assigned to individual cues during their combination are dictated by their respective variances (Ernst and Banks 2002), a mechanism referred to as “reliability-based cue weighting” (e.g. Hillis et al., 2004). This bias towards the more reliable single cue in a manner proportional to that cue’s reliability has also been demonstrated for inconsistent combined cue estimates such as visual texture and motion cues (Young et al. 1993) or multisensory cue combinations (Fetsch et al. 2012). However, whether inconsistent cue combinations result in increased reliability seems to depend on the degree of inconsistency and the individual cues’ reliabilities (Landy et al. 1995). The visual system derives these cue reliabilities from the variance of a cue’s distribution of inferences (Jacobs 2002), as well as from prior knowledge (Ernst and Bülthoff 2004) and visual context (Johnston et al. 1994).

Strikingly, despite the large body of research on cue combination, studies of shading and motion cue combinations are absent. The to our knowledge most closely related study focused on the 3-way interaction of motion parallax, disparity, and shading, but did not fully address the pairwise integration of shading and motion (Schiller et al., 2011). The principal goal of the present study was to probe the mechanisms by which monocular shading and motion cues are combined for depth perception, and investigate whether this occurs according to the predictions of MLE schemes. Human observers judged the curvature of 3D objects defined by shading, motion, as well as by consistent or inconsistent cue combinations, while small eye movements during fixation were measured. Stimulus presentations varied in visual field position to anticipate possible perceptual effects of eccentricity on performance (Previc, 1990).

## Methods

### Participants

Nine right-handed participants (5 male, 4 female) aged 24-29 years (mean = 27.5 years,  $SD = 1.81$ ) took part in four sessions spread over two days. All had normal or corrected-to-normal vision and no history of neurological disorders. Before the start of the experiment, written and oral informed consent was obtained in accordance with the Helsinki declaration. Approval for this study was granted by the Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

### Stimuli

Stimuli consisted of 3D ellipsoids (7.5 deg diameter) viewed with a gaze angle orthogonal to the flat surface that formed the background of the stimuli. Stimuli were defined either by shading, motion or combined shading and motion cues, presented on a light grey background (Fig. 1A). The x-y semi-principal axes of the ellipsoid formed the background plane, while the z-axis was orthogonal to the background and directed to the eyes of the participants. The x-y semi-principal axes always remained equal in length so that the shape's base appeared fully circular. The length of the z-axis was manipulated, and the resulting curvature changes were quantified on a normalized scale ranging from 0 to 10. A curvature of 10 describes a perfect half sphere protruding from the background by 3.75 deg towards the observer and a curvature of 0 describes a flat disk. In the experiment (Fig. 1E), the reference stimulus had 50% of the maximum curvature (curvature 5), whereas test stimuli varied between step 1 (10% of max curvature) and step 9 (90% of max curvature). Stimuli were created using MATLAB (R2009b, The MathWorks, Natick, Massachusetts, USA).

#### ***Shading-defined stimuli***

In shading stimuli (Fig. 1A, top row), the curvature of the ellipsoid was to be inferred from the luminance intensity distribution across the stimulus (average luminance: 38  $\text{cd}/\text{m}^2$ , range: 33 – 41  $\text{cd}/\text{m}^2$ ). To avoid participants using local cues to judge the curvature information, positions of the light sources were randomly varied between consecutive stimuli within each trial (top left, top right, bottom left, bottom right) (Fig. 1B). Thus, ellipsoids had one of 9 curvatures (manipulation of the length of the z-axis; 1 - 9) and were presented with one of the 4 light source position, resulting in 36 different shading stimuli.

#### ***Motion-defined stimuli***

The motion-defined ellipsoids (Fig. 1A, second row) were defined by a motion flow field ( $9.5 \times 9.5$  deg) on a light grey background. Within this surface, a  $42 \times 42$  grid of possible dot starting points was defined, of which 90% was occupied, so that the

flow field contained 1588 white dots ( $\sim 0.1$  deg diameter) moving coherently in an upward direction (from bottom to top of the flowfield). Dot life-time was 250 ms (5 frames at play rate of 20 Hz), and dot birth was distributed over frames so that in each frame approximately 20% of the dots died and were randomly reborn. Each dot was born within a circular region centred on a grid point, with this circular region defined by a radius equal to 50% of the (closest) distance between neighbouring grid points. Spatial overlap of dots was avoided. Dots were created on the background plane and were then projected onto the ellipsoid so that in the frontal perspective of the observer there was a fixed spacing between adjacent grid points. Moreover, dot size and shape was constant. Hence, dot density, size and shape did not provide a cue to depth perception. By contrast, each individual dot moved at a constant speed on the surface of the shape (4 deg/sec), leading to differences between dots' speed in the observer's frontal view. Thus, gradients in dot speed in the observer's frontal view were the only cue from which surface curvature could be derived. There were 9 different motion-defined stimuli (curvature steps 1-9), with for each trial a different movie drawn from 4 different random versions per curvature. Average luminance of the motion defined stimuli was ( $58 \text{ cd/m}^2$ ).

### ***Combined cue stimuli***

In combined cue stimuli, a motion-defined ellipsoidal stimulus was superimposed on a shading defined stimulus of equal size (average luminance:  $60 \text{ cd/m}^2$ , range:  $53\text{-}66 \text{ cd/m}^2$ ). As in the single shading stimuli, light source position was varied to avoid the use of local stimulus cues for curvature judgments (Fig. 1B). There were 36 'consistent' combined (CC) stimuli (9 curvature steps  $\times$  4 light sources) in which shading and motion described the same curvature (e.g. shading 4 paired with motion 4) (Fig. 1A, third row). Additionally, there were 4 inconsistent combined (IC) stimuli in which shading and motion cues described inconsistent curvatures (shading 5; motion 7  $\times$  4 light sources), which functioned as reference stimuli in inconsistent combined trials (Fig. 1A, bottom row; see also below "Task and trial structure" for details on test and reference). For each of the 36 curvature  $\times$  light source combinations of the consistent condition, and the 4 combinations for the inconsistent condition, a unique randomized version of the motion stimulus was superimposed to form either consistent or inconsistent combined stimuli.

### **Experimental design and task**

Experiments were conducted in a dimly lit room where stimuli were presented on a Samsung screen (size  $37.6 \times 30.2$  cm, refresh rate 60 Hz) at 57 cm viewing distance. Observers were seated in a comfortable chair with their heads stabilized by a chin and head rest and gave their responses through a standard keyboard.

### ***Task and trial structure***

In all sessions, each trial consisted of a reference stimulus (1.5 s duration) followed by a test stimulus of the same duration (method of constant stimuli). To minimize the possibility for misinterpretation of the status of the presented stimuli, refer-

ence and test stimuli were indicated to the observers by red and green fixation crosses, respectively. Reference and test stimuli were separated by a 200 ms inter-stimulus-interval (ISI), during which the fixation cross remained red until the test stimulus appeared. In each trial, participants were asked to compare the 3D curvature of the test stimulus to the curvature of the preceding reference stimulus. In a subsequent 1.3 s response period containing a white fixation cross they indicated by right-hand button press whether the test stimulus described a smaller or a larger radius than the reference stimulus (left arrow on the computer keyboard when the stimulus was 'flatter'; right arrow when it was 'rounder') (see Fig. 1D). No feedback was given to limit contributions of perceptual learning.

### ***Blocks, conditions and sessions***

The experiment was divided into 4 sessions of approximately 40 min (540 trials in total per session per participant) that were acquired on two separate testing days. On both testing days the two sessions were separated by a 15 min break. Each session was divided into blocks of maximally 10 min after which participants had the opportunity to take a break and restart the experiment by button-press. Prior to the start of each session, participants familiarized themselves with the stimuli and task in practice blocks lasting approximately 3 min.

#### *Single cue sessions*

Single cue shading and motion stimuli were tested separately in two sessions on the first testing day. In each single cue trial the reference described a curvature of '5' followed by a test stimulus (curvature 1 - 9).

#### *Combined consistent and inconsistent sessions*

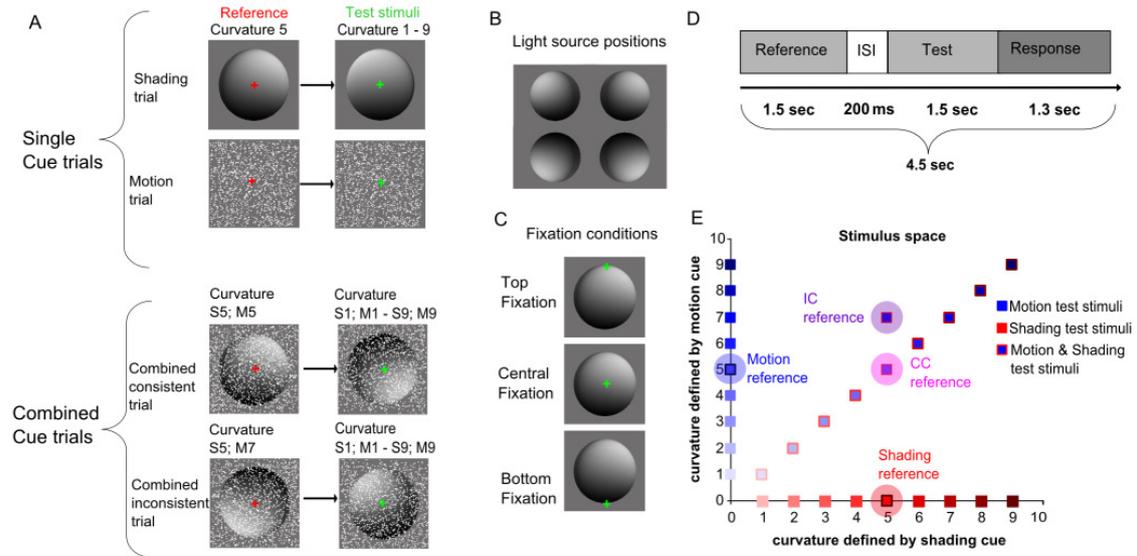
Combined consistent and inconsistent cues were tested in two sessions on the second testing day. In the consistent combined cue trials, the reference stimulus described a curvature of 5 (i.e. shading 5; motion 5) followed by combined consistent test stimuli ranging from curvature 1 to 9 (e.g. shading 2; motion 2). In inconsistent combined cue trials, the reference stimulus consisted of inconsistent curvature information, i.e. shading- and motion-defined curvatures were conflicting (shading 5; motion 7) whereas the test stimuli were consistent shading/motion pairs (curvatures 1 - 9). Thus, test stimuli were always consistent while the reference stimulus could either be consistent (shading 5; motion 5) or inconsistent (shading 5; motion 7). The terms consistent or inconsistent condition refer, respectively, to trials in which the reference stimulus is made of consistent or inconsistent cues. Consistent and inconsistent trials were presented in random order throughout the session. When questioned after completing the task, participants reported to have been unaware of stimulus inconsistencies throughout the experiment.

### *Fixation conditions*

In addition to the two single cue and the two combined cue conditions, there were three fixation conditions (Fig. 1C). In the ‘top’ condition, the fixation cross was at the top edge of the ellipsoid (3.75 deg above the centre of the stimulus); in the ‘bottom’ condition it was at the bottom edge of the stimulus; and in the ‘central’ condition it was placed centrally. The position of the fixation cross was always the same in reference, test and response period and was varied in blocks of ~ 3 min. At the beginning of each fixation block a fixation cross at the relevant position was shown for 1000 ms to indicate the fixation position to the participant. For each of the 4 trial types (single shading, single motion, consistent combined, inconsistent combined), 3 fixation conditions (top, central, bottom) and 9 test curvatures (108 conditions), 20 trials per condition were collected. Figure 1E summarizes the conditions and stimulus combinations.

### *Control experiment*

In 5 of the 9 original participants, the data were replicated for single as well as consistent and inconsistent combined cues. The primary goal of this control experiment was to show that the trial-by-trial randomization of consistent and inconsistent conditions had not interfered with the visual system’s reliability estimations and created a context that prevented maximum-likelihood estimation. The replication was limited to the central fixation condition. On the first testing day, single shading and single motion cue conditions were tested. For each cue, there was one run, each split into two 7-min blocks (order of runs randomized over participants). On the second testing day, combined cue conditions were tested in a total of two runs. In one run (randomized condition), trials of the consistent and of the inconsistent condition were presented in randomized order (as in main experiment). In the other run (non-randomized condition), blocks contained either only consistent trials or only inconsistent trials. Each run was subdivided in four 7-min blocks of trials, and the order of runs was randomized over participants. In each run we collected 20 trials per curvature condition. The replicated data were obtained with a new set of 76 stimulus versions with randomized motion information (36 for the single motion cue, and 40 for the combined condition; see *Single cue stimuli* and *Combined cue stimuli*). Slope data obtained from fitted curves based on this new dataset were subjected to repeated-measures ANOVA with the factors stimulus presentation (randomized, non-randomized) and cue type (consistent, inconsistent). Results revealed neither significant main effects nor interactions ( $p > .05$ ). Hence, the absence of differences in reliability between data obtained in randomized and non-randomized control conditions confirm that the randomization of consistent and inconsistent conditions did not contribute to the reported patterns in our data.



**Fig. 1. Experimental design.** **A.** Illustration of the 4 stimulus conditions. In each trial, reference and test 3D ellipsoidal shapes were defined by single shading, single motion, or combined cues. Stimuli were made to suggest that ellipsoidal shapes protruded from a frontal plane towards the observer, along a gaze angle orthogonal to the plane. For the single shading cues, shapes were defined by luminance gradients whereas in the single motion condition shape was rendered by white dots moving on a grey background. In the combined conditions, shape was described by white dots sliding over the shading-defined shape. The dot pattern in a single frame (depicted) of the motion and combined stimuli did not contain shape information (see ‘Methods’ for details on stimuli). Participants were required to judge whether the test stimuli contained more (greater convexity) or less (smaller convexity) curvature than the reference stimulus. **B.** Variations of light source position. In single shading and in combined cue trials (not shown), the light source varied within a trial between 4 possible positions to prevent local cues from contributing to shape perception. **C.** Fixation conditions. The fixation cross was either located at the top, the center or the bottom of the shape. **D.** Stimulus timing of one trial. A reference stimulus was followed by a brief inter-stimulus interval (ISI), a test stimulus, and a response window. **E.** Stimulus space showing the reference and test stimuli used in the 4 cue conditions of the present experiment. Small squares in red (x-axis) and blue (y-axis) show stimulus space positions of single shading and motion cues, respectively. Saturation increase symbolizes the curvature increase in equal (linear) steps from almost flat (z-distance 1) to almost maximally convex (z-distance 9). Curvature is quantified by the distance from the point of the shape closest to the observer to the background (z-distance), with z-distance 0 corresponding to a flat surface and z-distance 10 to a perfectly convex half sphere protruding from the flat background. Symbols on the diagonal show test stimuli made of matching shading and motion cues. The Consistent Combined (CC) reference stimulus consists of shading (S) and motion (M) cues at curvature 5. The Inconsistent Combined (IC) reference stimulus combines curvature 5 from the shading cue with curvature 7 from the motion cue (see ‘Methods’ for details).

## Analysis

### *Analysis of curvature discrimination performance*

Psychophysical data were analysed using MATLAB (R2010b, The MathWorks, Natick, Massachusetts, USA) and SPSS statistical software (IBM SPSS Statistics, version 19). After exclusion of invalid trials (see “Eyetracking”), we used logistic regression to fit psychometric response functions to the psychophysical data, plotting the proportion ‘larger curvature responses’ as a function of curvature. We

employed a curve-fitting procedure (using Matlab's "fit" function) using the logistic function given by Equation 1:

$$y = \frac{a}{1 + e^{-\frac{(x-d)}{c}}} + b \quad (1)$$

where  $a$  provides a measure of the curve amplitude,  $b$  corresponds to the y-axis location of the lower asymptote,  $c$  indicates the slope and  $d$  reflects the location of the midpoint of the curve on the x-axis. The nonlinear least-squares parameter estimation was constrained as follows:  $0 \leq a \leq 1$ ;  $0.1 \leq b \leq 1$ ;  $0.2 \leq c \leq 10$ ;  $0 \leq d \leq 9$ . These liberal parameter settings were chosen to achieve a good fit ( $r^2 > .85$  in all individually fitted curves and  $r^2 > .95$  in average curves, Fig. 2).

Subsequently, we determined 3 parameters from these lines of best fit that describe different aspects of psychometric performance. First, as an estimate of bias, we calculated points of subjective equality (PSE) (curvature in the test stimulus judged equal to curvature in the reference stimulus; curvature corresponding to the 0.5 point on the y-axis). Second, to capture situations in which the floor and top of the psychometric functions did not reach 0 and 1, we determined curve amplitudes by calculating the distance between the lower and the upper bound of the curve. This measure can be interpreted as a reflection of the noise level of 3D perception at the two extremes of the range of curvatures used (anchor points). Thirdly, we calculated the slopes of the curves at the mid-point of the curve (parameter  $d$ ), which was defined by equation 2.

$$slope = \frac{a}{4c} \quad (2),$$

Slope provides an accurate measure of discrimination precision (sensitivity), and can be seen as a measure of reliability, which is the terminology we have chosen to use here. Differences between conditions in PSE, amplitude and slope measurements were tested using repeated measures ANOVAs with within-subject factors curvature (1 – 9), fixation condition (top, central, bottom) and cue type. We conducted four a priori planned analyses. The first analysis pertained to the differences in 3D shape judgment between cues defined by shading and motion (single cues) while the second and third analyses investigated differences between single cues and combined consistent cues. The fourth analysis compared consistent and inconsistent combined cue trials.

### **Cue weight estimation**

Based on the psychometric functions for each condition we intended to quantify the contributions of each single cue to the combined estimate. This was done in two different ways. First, we determined the weighted average of single-cue psychometric functions best fitting (smallest squared error) the observed con-

sistent combined cue psychometric function. To that goal, we used an iterative procedure in which the weights of shading and motion were manipulated to generate a predicted combined cue psychometric distribution. The set of weights producing the smallest deviation of the predicted combined cue curve from the empirical combined cue curve (as estimated from squared error) is considered to give the shading and motion weights of the weighted averaging underlying the consistent combined psychometric curve.

Secondly, we estimated cue weights for shading and motion cues based on PSEs in consistent conditions (averaged across fixation conditions). Following a linear model of cue integration we postulated that the PSE in the consistent combined condition should be a weighted sum of single motion and shading PSEs, where the sum of the weights add up to 1 ( $w_m + w_s = 1$ )

$$PSE_{CC} = w_m PSE_m + w_s PSE_s \quad (3)$$

in which case the motion weight is given by

$$W_m = \frac{PSE_{CC} - PSE_s}{PSE_m - PSE_s} \quad (4)$$

In the case of the inconsistent conditions, the consistent combined PSE is derived from the single shading PSE and a single motion PSE shifted to the right by 2 x-axis units, corresponding to an estimate of PSEs obtained in a single motion condition with a reference stimulus of 7.

$$PSE_{IC} = w_m (PSE_m + 2) + w_s PSE_s \quad (5)$$

Subsequently, we tested on a single participant basis whether the motion cue weights estimated from the PSEs in consistent combined conditions were predictive of the motion cue weights obtained from PSEs in the inconsistent combined condition.

Finally, to test the prediction of the MLE rule concerning a relationship between the sensitivity/reliability of each single cue and its cue weight in a combined cue estimate, we performed separate linear regression analyses for the predictors “slope motion” and “slope shading” and tested whether these measures significantly predicted cue weights based on PSE or based on the iterative procedure in the consistent combined condition.

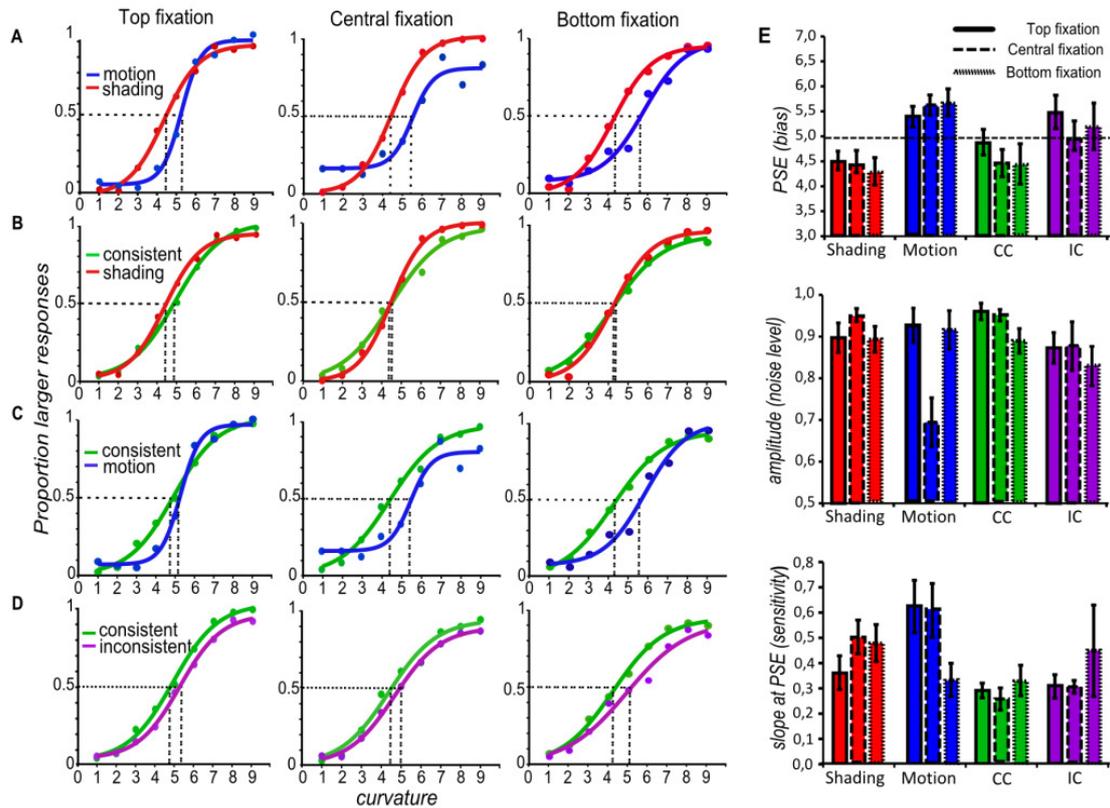
### ***Analysis of response times***

Trials with response times that were 1.5 *SD* above or below the mean were regarded as outliers and were removed from response time analysis. Average response time data for each condition and participant were subjected to a repeated measures ANOVA with factors curvature (1 - 9), fixation condition (top, central,

bottom) and cue type (single shading, single motion, consistent combined, inconsistent combined). When applicable, p-values were corrected for non-sphericity using the Greenhouse-Geisser correction. Response time cue weights were calculated by considering average RTs of each cue condition pooled over fixation positions.

### ***Eyetracking***

Eye position was monitored with an infrared eye camera (60 Hz sampling rate, Viewpoint Eye Tracker v.2.8.3, Arrington Research) and eye movements were continuously recorded using the Viewpoint eyetracking software. Using Matlab R2010b, eye movement data of each trial data were epoched considering the time window until the end of the test stimulus or until a response was given (for cases in which responses were given prior to the end of the test stimulus). Trials containing eye blinks and eye deviations exceeding 2 degrees in x or y direction relative to fixation were discarded from the analysis, yielding a subset of accurately fixated trials for all psychophysical data analysis in this paper. In one additional analysis in order to investigate effects of fixation accuracy on cue weighting, we also considered trials that deviated up to 3 degrees from fixation. Eye movements within short distances from fixation are considered “fixation saccades” (Gowen et al. 2007) comprising both micro-saccades and saccadic intrusions during fixation. Fixation accuracy within the specified range of 2 degrees was quantified as the minimum ellipsoid area (MEA) encompassing a cluster of all gaze points (x-y coordinates in world space sampled at 60 Hz) during our specified time bin (test stimulus duration or until a response was given). In each participant the MEA was computed in each trial and was then averaged over all trials of the respective conditions. This measure represents the magnitude of fixation errors in squared visual degrees (Nachmias 1959; Sun and Freund 2004).



**Fig. 2 Comparisons of psychometric response functions of single and combined cue conditions.** **A** Single shading and single motion curves (blue and red lines). **B** Single shading and consistent combined curves (red and green lines). **C** Single motion and consistent combined curves (blue and green lines). **D** Consistent and inconsistent combined curves (green and purple lines). The proportion of responses in which participants perceived the curvature of the test stimulus as larger than the reference (“proportion larger responses”; y-axis) is plotted as a function of curvature (x-axis). Dashed lines indicate the PSEs. **E** Summary of the 3 parameters estimated from psychometric curves. Top: PSE to measure bias; Middle: total amplitude parameter to measure noise; Bottom: slope at the midpoint of the curve as an indication of sensitivity/reliability (larger values indicate a steeper slope). Bars with solid, dashed, and dotted outlines refer to data from top, central, and bottom fixation conditions. Colors of the bars refer to the different stimulus conditions. (red = shading, blue = motion, green = consistent combined, purple = inconsistent combined). The dashed line in the upper panel indicates the physical midpoint where reference and test stimuli are equal. Error bars represent the standard error of the mean (SEM).

## Results

### Psychophysics

For statistical analysis, logistic regression was used to calculate lines of best fit using the decision value per curvature level (proportion responses test curvature > proportion responses reference curvature). Based on our interest in differences between single cue and combined cue performance parameters, we conducted 4 a priori planned 2 (cue type) × 3 (fixation top, central, bottom) repeated-measures ANOVAs, pertaining to comparisons between single cue conditions (shading vs

motion), single cue and combined conditions (shading vs consistent combined; motion vs consistent combined) as well as consistent and inconsistent combined conditions. Parameters of interest for the ANOVAs were the perceptual bias associated with an estimate (reflected by the PSE; corresponding to the point where curvatures of the test and reference were judged equal), baseline noise level of the estimate (amplitude; difference between upper and lower bound of the curve) and sensitivity for curvature differences (reliability; slope of the curve at point  $d$ ).

### **Comparing single shading and single motion cues**

In a first ANOVA of **PSE** data we compared the effectiveness of single shading and single motion cues for curvature perception (Fig. 2A). We found a significant effect of cue ( $F_{(1,8)} = 19.142$ ,  $p = .002$ ) but no significant effect of fixation position or interaction term ( $p > .05$ ). The main effect of cue type is shown by a rightward shift of the psychometric function for the motion relative to the shading condition (Fig. 2A). This shows that participants experienced motion-defined mid-level test curvatures (steps 4-6) as smaller than when they were defined by shading. A comparison of shading and motion cue PSEs with the physically defined (“real”) step 5 point revealed a bias towards *overestimation* of mid-level curvatures in the shading cues ( $F_{(1,8)} = 8.946$ ,  $p = .017$ ) and a bias towards *underestimation* of mid-level curvatures in the motion cues ( $F_{(1,8)} = 7.046$ ,  $p = .029$ ) in all fixation conditions (no sign. effects of fixation condition or interaction term;  $p > 0.05$ ) (see Fig. 2E, upper panel). In an ANOVA of curve **amplitude** data, we also found a significant main effect of cue type ( $F_{(1,8)} = 9.900$ ,  $p = .014$ ) reflecting a greater amplitude for shading, no significant main effect of fixation position ( $p > .05$ ) but a significant interaction term ( $F_{(2,16)} = 10.878$ ,  $p = .001$ ). Post-hoc ANOVAs per fixation position indicated an effect of cue type only in the central fixation position ( $F_{(1,8)} = 26.135$ ,  $p = .001$ ), witnessed by a decreased amplitude in the single motion condition. This effect was not significant in the top and bottom fixation positions ( $p > .05$ ). An analogous ANOVA with the sensitivity measure (**slope**) revealed no significant main effects of condition or fixation position ( $p > .05$ ) but a (barely) significant interaction term ( $F_{(2,16)} = 4.679$ ,  $p = .049$ ). Testing for differences in slope for each fixation position separately revealed a marginally significant main effect of slope in the top fixation position ( $F_{(1,8)} = 4.139$ ,  $p = .076$ ), pertaining to higher sensitivity for the motion compared to the shading cue (Fig. 2A, left panel), but no differences in the central and bottom fixation positions.

Taken together, these findings suggest that although both cues were about equally sensitive to curvature differences, motion cue estimates were biased towards smaller and shading cue estimates were biased towards larger mid-level curvature judgments, compared to the physically defined point at which reference and test were equal. However, since for the motion cue noise level (estimated by curve amplitude) varied with fixation condition, as evidenced by smaller amplitudes in the central fixation condition, the shading cue (performance unaffected by fixation position) may be considered a more robust estimate of 3D curvature than motion.

### **Comparing single cues and consistent combined cues**

Our second question pertained to performance differences between single cue conditions (shading and motion alone) and combined consistent conditions (Fig. 2B,C).

The comparison of **PSE** data between single shading and consistent combined conditions (Fig. 2B) did not show significant main effects of cue type or fixation position and no significant interaction term ( $p > .05$ ). Interestingly a comparison between PSEs in the consistent condition and the physically defined midpoint (curvature 5) revealed no significant difference ( $F_{(1,8)} = 2.148, p = .181$ ) and no main effect of fixation or interactions ( $p > .05$ ), showing a veridical (unbiased) estimate of curvature from combined cues. Analysis of curve **amplitude** data did not show significant main effects or interactions. Interestingly, **slope** analysis indicated higher sensitivity of the single shading cue compared to the consistent combined cue (sign. main effect of cue type:  $F_{(1,8)} = 21.173, p = .002$ ), irrespective of fixation position (no main effect of fixation position and no significant interaction,  $p > .05$ ).

The second comparison, a repeated measures ANOVA using **PSE** data from motion and combined consistent conditions (Fig. 2C), revealed no main effect of fixation position ( $p > 0.05$ ) but a significant main effect of cue type ( $F_{(1,8)} = 14.553, p = .005$ ), which however depended on fixation position (interaction fixation position  $\times$  cue type:  $F_{(2,16)} = 4.367, p = .031$ ). Separate post-hoc ANOVAs revealed significant main effects of cue type only in the central and bottom fixation position ( $F_{(1,8)} = 24.977, p = .001$  and  $F_{(1,8)} = 9.503, p = .015$ , respectively), corresponding to a rightward shift of the curve for the motion compared to the consistent combined stimuli. An analysis using curve **amplitude** revealed significant main effects of cue type ( $F_{(1,8)} = 12.563, p = .008$ ) and fixation position ( $F_{(2,16)} = 4.281, p = .032$ ) as well as a significant interaction term ( $F_{(2,16)} = 10.266, p = .001$ ). These effects were followed up in separate ANOVAs per fixation position and showed to correspond to a decreased amplitude of the motion curve compared to the combined consistent curve only in the central fixation position, but not in the top and bottom fixation position conditions (Fig. 2C). A corresponding ANOVA using the **slope** parameter revealed increased sensitivity of the single motion cue compared to the consistent combined cue (main effect of cue type:  $F_{(1,8)} = 15.699, p = .004$ ) that however depended on fixation position (interaction  $F_{(2,16)} = 3.699, p = .048$ ). Fixation position was not significant ( $p > .05$ ). Follow-up analysis for each fixation position revealed higher sensitivity for the motion than the consistent combined cue in the upper ( $F_{(1,8)} = 8.346, p = .02$ ) and in the central fixation ( $F_{(1,8)} = 11.191, p = .010$ ), but not in the bottom fixation condition ( $p > .05$ ).

In sum, consistent combined cues provided veridical curvature estimates whereas PSEs of both shading and motion cues differed significantly from the physically defined midpoint. Similar to the comparison between single shading and motion conditions (compare Fig. 2A), the motion cue compared to the consistent combined cue showed a bias towards underestimation of mid-level curvatures in the central and bottom fixation condition (shift in PSE) as well as a decreased amplitude (higher noise level) when participants fixated centrally. Shading and consistent combined conditions did not differ in terms of PSE and amplitude.

Interestingly, sensitivity (slope) was greater in *both* single cue conditions than in the consistent combined condition.

### **Comparing consistent combined and inconsistent combined cues**

In a third analysis, we compared performance in consistent combined and inconsistent combined conditions (Fig. 2D). A repeated-measures ANOVA of *PSE* data showed a reliable difference between cue types ( $F_{(1,8)} = 20.563$ ,  $p = .002$ ) while fixation position and the interaction term did not reach significance ( $p > .05$ ). Note that the difference between consistent and inconsistent conditions corresponded to a lateral shift to the right in the psychometric function for inconsistent relative to consistent conditions (see Fig. 2D). While PSEs differed significantly, the amplitudes and slopes of the underlying psychometric functions did not (all  $p > 0.05$  in corresponding ANOVAs on curve amplitudes and slopes). These results show that in inconsistent combined conditions, participants combined curvature information from both shading and motion cues. However, consistent with previous analysis showing no differences between consistent combined cues and shading, the shift to the right was relatively small considering the 2 curvature steps by which motion-defined curvature exceeded shading-defined curvature in the reference stimulus (see Fig. 2D). This suggests that also in inconsistent conditions information conveyed by the shading cue dominated the combined cue percept while the influence of the motion cue was comparatively small.

### **Calculating cue weights**

The above reported statistics showed decreased sensitivity (slope) in combined conditions compared to both single cue conditions and hence argue against an optimal combination of shading and motion cues as predicted by MLE. Moreover, the respective single cue sensitivities/reliabilities did not differ, which would - according to the MLE rule - lead to a prediction of approximately equal weighting of the cues in a combined condition. The preceding analysis does not support this prediction.

Instead, the data shown in Figure 2 are suggestive of a weighted averaging of cues, with psychometric curves for the consistent combined cue falling in between the psychometric curve for the consistent combined condition. Here we used two procedures to derive weights of single shading and motion cues to the combined cue depth estimates. First, we applied an iterative weighted averaging approach to determine the weights for single cue curves that led to an average curve with the best fit to the empirically observed combined curve (see Methods). The resulting weight estimates were based on the shape of the entire curve, thus reflecting aspects of sensitivity (slope), noise (amplitude) and bias (PSE). Based on group average curves, estimated relative cue weights seemed to depend on fixation position with the motion cue being weighted more heavily in the central fixation position ( $w_m = .37$ ) compared to the top ( $w_m = .09$ ) and bottom fixation position ( $w_m = .171$ ) (recall that  $w_s = 1 - w_m$ ; for simplicity we here only report motion weights). However, when fitting curves and deriving cue weights on an individual

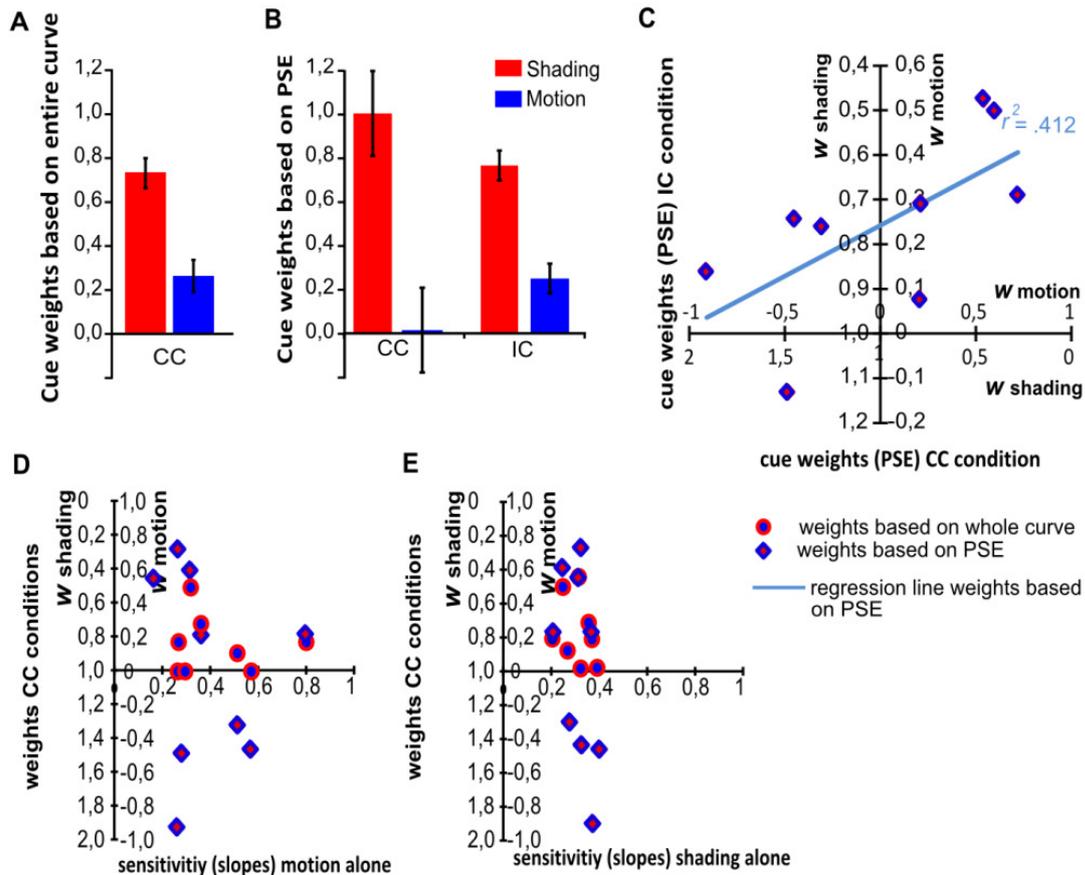
participant level, in a repeated measures ANOVA these differences between fixation positions were not statistically significant ( $p > .05$ ). Therefore data from the 3 fixation positions were pooled for further analysis resulting in cue weights of .26 for motion and .74 for shading in the pooled group average curves (Fig. 3A).

Secondly, we determined single shading and motion cue weights based on the PSEs of individual participants in consistent and inconsistent conditions (see Methods). Average motion cue weights were 0.01 and 0.25, respectively in the consistent and inconsistent condition (Fig. 3B). Thus, these analyses confirm a strong dominance of the shading cue.

In a subsequent analysis, we tested whether single cue weights derived from PSEs in the consistent condition predicted single cue weights derived from expected PSEs in inconsistent combined conditions (Fig. 3C). This analysis showed a marginally significant positive correlation ( $r^2 = .412$ ,  $\beta = .225$ ,  $F_{(1,7)} = 4.914$ ,  $p = .062$ ), suggesting that there was a trend for cue weights to be roughly similar in consistent and inconsistent conditions on an individual participant basis. Note that the cue weights are negative in single participants in case the value of the PSE in the consistent combined condition is smaller than the PSE in the shading condition *or* if the PSE in the motion condition is smaller than the PSE in the shading condition (see Methods; equations 4 and 5).

As described above, the MLE rule predicts a relationship between cue weights in combined estimates and single cue reliabilities. To explicitly test this prediction, we performed regression analyses in which we tested whether the sensitivity/reliability (slope) of single motion (Fig. 3D) and single shading cues (Fig. 3E) correlated with cue weights in the consistent combined condition. These relationships revealed non-significant both using cue weights estimated from PSEs in consistent combined conditions (for motion:  $r^2 = .021$ ;  $\beta = -4.19$ ,  $F_{(1,7)} = .154$ ;  $p = .707$ ; for shading:  $r^2 = .243$ ;  $\beta = 4.577$ ,  $F_{(1,7)} = 2.244$ ;  $p = .178$ ; Fig 3D and E, blue outlined diamonds) as well as using weights derived from fitting of whole-curve weighted averages (for motion:  $r^2 = .12$ ;  $\beta = -.355$ ,  $F_{(1,7)} = .951$ ;  $p = .362$ ; for shading:  $r^2 = .105$ ;  $\beta = 1.078$ ,  $F_{(1,7)} = .820$ ;  $p = .395$  (Fig. 3D and E, red outlined disks). Hence, these results demonstrate no relationship between single cue reliabilities and their respective weighting in a combined percept.

In sum, our data show that irrespective of the manner in which cue weights are determined, the shading cue was dominant over the motion cue. Single cue weights derived from PSEs in consistent combined conditions tended to correspond to predicted cue weights in inconsistent combined conditions. However, single cue weighting in the combined consistent condition was unrelated to respective single cue reliabilities, which did not differ between single shading and single motion cues. This argues against MLE as a mechanism for shading/motion cue combination in the present experiment which would have predicted equal cue weights in combined conditions.



**Fig. 3 Cue weight estimation** **A** An iterative procedure taking the shapes of the entire shading, motion and consistent combined curves into account (averaged across fixation conditions) yielded larger cue weights for shading than for motion **B** Cue weights for shading and motion estimated from PSEs in CC and IC conditions **C** The shading and motion cue weights derived from PSEs in the consistent condition (marginally) significantly predicted their respective weights in the inconsistent condition (light blue regression line). **D&E** Neither the single motion nor the single shading cues' sensitivities/reliabilities (slopes) (x-axes) were predictive of their respective cue weights in the CC condition (y-axes). This was the case for cue weights based on PSE and based on the entire curve. Error bars represent the standard error of the mean. Red outlined disks represent the weights based on the whole curve and blue outlined diamonds represent the weights based on PSEs.

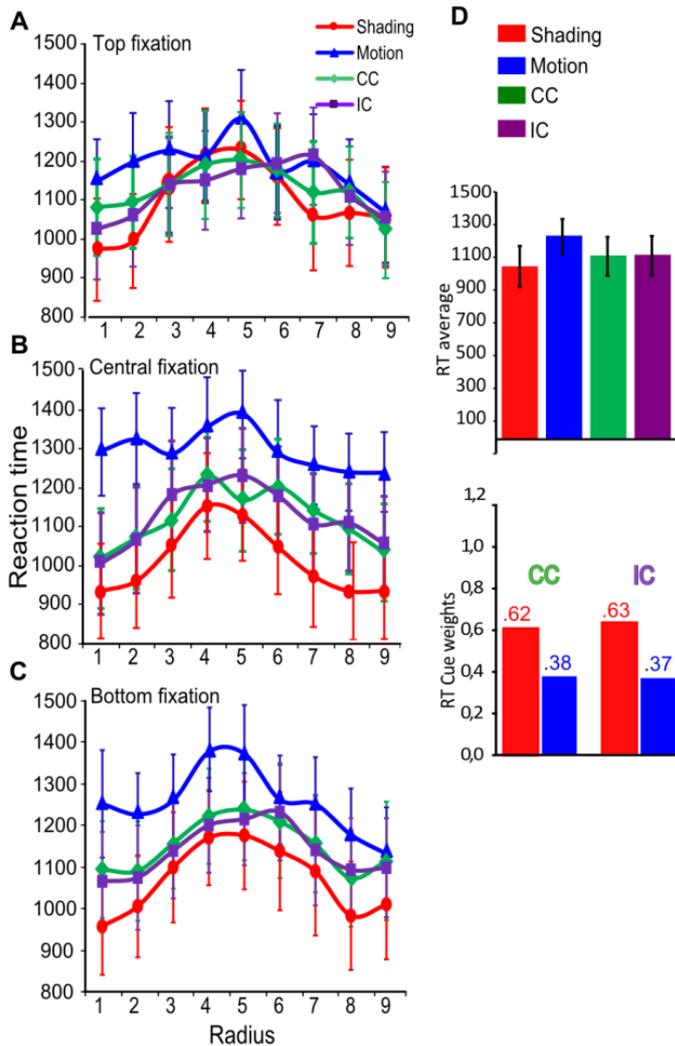
## Response times

A 4 cue type (S, M, CC, IC)  $\times$  3 fixation (top, central, bottom)  $\times$  9 curvature levels (1 - 9) repeated measures ANOVA on response times revealed significant main effects of cue type ( $F_{(3,24)} = 10.026, p = .002$ ) and curvature ( $F_{(8,64)} = 10.510, p = .002$ ) but not fixation position ( $p > .05$ ). However, the main effect of cue type depended on fixation position (interaction: cue type  $\times$  fixation position:  $F_{(6,48)} = 6.859, p = .001$ ) and curvature (interaction: cue type  $\times$  curvature  $F_{(24,192)} = 1.888, p = .010$ ). Hence separate post-hoc analyses per fixation position with within-subject factors cue type and curvature were conducted.

In the top fixation condition, cue type did not have a significant effect on response times ( $p > .05$ ), but there was a significant main effect of curvature ( $F_{(8,64)} = 8.042$ ,  $p < .001$ ), indicating that participants' response times increased with task difficulty (increasing similarity between test and reference curvature) (Fig. 4A). The interaction term missed significance ( $p > .05$ ). In the central fixation condition, both cue type and curvature showed significant main effects ( $F_{(3,24)} = 16.767$ ,  $p < .001$  and  $F_{(8,64)} = 6.750$ ,  $p = .006$ , respectively), but no interaction ( $p > .05$ ). The effect of cue type was due to longest response times for single motion stimuli, intermediate response times for combined stimuli (consistent and inconsistent) and shortest response times for single shading stimuli, while the main effect of curvature reflected differences in task difficulty among conditions (Fig. 4B). Similarly, in the bottom fixation condition, both cue type and curvature ( $F_{(3,24)} = 8.586$ ,  $p = .006$  and  $F_{(8,64)} = 6.597$ ,  $p = .003$ , respectively) had a significant effect on response time and did not interact ( $p > .05$ ). Again, the effect of cue type pertained to fastest response times for shading stimuli, intermediate response times for combined stimuli (consistent & inconsistent) and longest response times for single motion stimuli while the effect of curvature indicated longer response times for difficult conditions (Fig. 4C). In sum, response times were faster for trials in which test condition curvatures were dissimilar from the reference compared to trials in which test stimuli curvatures were similar or equal to the reference. Furthermore, in the central and bottom fixation condition, participants responded fastest for single shading stimuli, intermediately fast for combined stimuli and slowest for the single motion stimuli.

Additionally, to investigate whether RT on average differed significantly between fixation positions, we averaged across curvatures and computed repeated-measures ANOVAs with the factor "fixation position" for each cue separately. Interestingly, while combined (consistent & inconsistent) and shading cue RTs showed no dependence on fixation position ( $p > .05$ ), the single motion condition showed a significant main effect of fixation position ( $F_{(2,16)} = 9.708$ ,  $p = .002$ ). RT were faster in top compared to the central fixation ( $p = .017$ ) and also marginally faster in top compared to the bottom fixation condition ( $p = .056$ ) ( $p$ -values corrected for multiple comparisons). Interestingly, these faster RT in the top fixation condition may also be related to the high sensitivity (steep slope) of the motion cue in the discrimination performance data in this condition (Fig. 2A). Similarly, the slow RTs in the central fixation condition are in line with the increased noise level associated with performance in this condition (i.e. amplitude reduction of the curve).

To estimate cue weights based on RT we pooled over all curvature levels and fixation conditions and used the average RTs for each condition (Fig. 4D, upper panel) to derive relative shading and motion cue weights (Fig. 4D, lower panel). Consistent with previous analyses motion was weighted less strongly than shading ( $w_m = .38$  and  $w_m = .37$ , respectively in the consistent and inconsistent combined conditions).



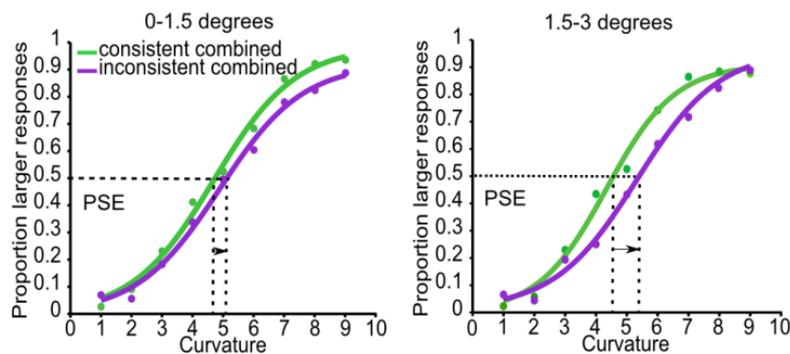
**Fig. 4 Response time data.** A-C Response times varied with task difficulty, being shorter for trials in which test stimuli were dissimilar from the reference and longer in trials in which test stimuli were similar or equal to the reference. Cue type also affected response times significantly with participants responding fastest for single shading stimuli, intermediately fast for combined stimuli and slowest for the single motion stimuli (B-C). **D** Average RT data pooled over all curvature and fixation conditions (top bar graph) from which cue weights for shading and motion were derived (lower bar graph). In accordance with the discrimination performance data, both in the consistent and inconsistent combined conditions, the shading cue was weighted more strongly than the motion cue. Error bars represent the standard error of the mean.

## Eyetracking

Although we already restricted our analysis to trials within 2 visual degrees from fixation, it remains possible that differences in fixation accuracy within this window contributed to performance differences among conditions, which would interfere with our interpretation of performance differences as reflecting perceptual differences. Therefore we computed MEA as a measure of fixation accuracy (see Methods) for each condition pooled across all curvatures. A repeated-measures ANOVA with the factors cue type (S, M, CC, IC) and fixation position revealed no significant

main effects or interactions ( $p < .05$ ), confirming no confounding effect of differences in fixation performance between conditions.

In a final analysis, we investigated the impact of eye movements on respective cue weighting by comparing the difference in PSE between consistent combined and inconsistent combined conditions in two subsets of the trials used throughout the whole analysis, i.e. trials fulfilling a strict fixation criterion (0-1 degrees) and trials meeting a moderate fixation criterion (1-2 degrees). To ensure a sufficient number of trials in each fixation band, trials from all fixation position conditions (top, central, bottom) were pooled before computing a repeated-measures ANOVA with within-subject factors cue type (consistent, inconsistent) and fixation performance (strict, moderate). Here, despite showing a trend for more weight assigned to motion in less strict fixation criteria trials, the interaction term did not reach significance ( $p > .05$ ). Nevertheless, a more pronounced effect of eye-movements might have been found had a larger range of eye movements been taken into account. To that goal we performed the same analysis as above with trials divided into fixation bands of (0-1.5 degrees) and (1.5-3 degrees). Results revealed a significant main effect of cue type ( $F_{(1,8)} = 34.607$ ,  $p < .001$ ), no significant main effect of fixation performance ( $F_{(1,8)} = .027$ ,  $p > .05$ ) but a significant interaction effect ( $F_{(1,8)} = 6.124$ ,  $p = .038$ ). Separate t-tests per fixation band revealed a smaller difference between PSEs in consistent and inconsistent stimuli for trials in the strict (0-1.5 degrees) fixation band ( $t_{(8)} = 2.620$ ,  $p = .031$ ) and a larger difference in trials in the moderate (1.5-3 degrees) fixation band (i.e. larger rightward shift in the curve for inconsistent compared to consistent conditions,  $t_{(8)} = 5.743$ ,  $p < .001$ ) (see Fig. 5). Hence, the magnitude of the motion cue weight in the inconsistent combined condition was dependent on eye movements. In particular, our data indicate that the weight of the motion cue increased when small eye movements around the fixation cross were tolerated.



**Fig. 5 Cue combination depends on eye movements. A** When participants fixated well (errors within 0-1.5 degrees from the fixation cross), the rightward shift of the curve was smaller than in a subset of trials with moderate fixation errors (during which participants fixated within a band of 1.5-3 degrees from the fixation cross), indicating an increased motion cue weight when small eye movements are tolerated. Dotted lines indicate the PSEs in the consistent and inconsistent combined conditions.

## Discussion

Our results indicate that shading and motion cues did not combine according to MLE during 3D curvature discriminations. Below, we discuss the factors that may have contributed to these findings.

### Reliability-based cue combination according to MLE: stimulus-specific constraints

Shading and motion cues, in the conditions of our experiments, did not combine in a statistically optimal fashion. Instead, curvature judgments in combined cue stimuli were less reliable than curvature judgments based on single cues. These findings are indicative of a form of suppressive competition. Hillis et al. (2002) pointed out that to benefit from MLE, cues need to be well correlated. However, in the present study, we found reaction times to be significantly longer for single motion than for single shading cues. While this is in line with a previous study reporting a similar delay for structure-from-motion (SFM) processing compared to static stimuli (Jiang et al. 2008), this may suggest that suboptimal cue integration in the present task was caused by the poor temporal correlation of depth-from-motion and depth-from-shading processing. However, RTs did not differ between shading and motion in the top fixation condition and yet, cue combined estimates were less reliable than single cue estimates. Therefore, a processing asynchrony may not solely account for the results reported here.

### Perceptual biases: MLE and alternative theoretical accounts

Single cue sensitivity/reliability could not account for the significantly larger weight of the shading cue in the combined percept, running counter the predictions from MLE. Interestingly, we found that while shading-defined curvature estimates were biased towards *overestimation* of mid-level curvatures, consistent with a previously reported bias towards a convex interpretation of shading-defined objects (Langer and Bulthoff 2001; Gerardin 2007), at the same time, motion-based curvature estimates were biased towards *underestimation*. Consistent combined estimates were more veridical, however *less sensitive/reliable* than single cues. The perceptual biases in single cue-based curvature estimation imply diverging estimates of curvature in objects defined simultaneously by both cues, which may have contributed to the observed lack of integration. Hence, although in the consistent combined condition the cues were physically matched, perceptually they were inconsistent and might therefore not have resulted in increased sensitivity (e.g. Ban et al., 2012). Specifically, in the case of small conflicts between multiple cues defining an object, previous studies have shown that perception reflects a weighted combination of cues (Johnston et al. 1994; Welchman et al. 2005), whereas large conflicts may result in single cue dominance (Rock and Victor 1964). Interestingly, here, the single cue biases may have caused large perceptual cue conflict in the *physically consistent* combined condition, predicting a winner-take-all (WTA) competition

between single cues (Battaglia et al. 2003) that may have resulted in the large weight of the shading cue. This implies that the *physically inconsistent* condition would have resulted in smaller perceptual conflict, assuming that the physically larger motion-defined curvature partially compensated for the perceptual bias. Although this idea is supported by more balanced cue weighting (although still in favour of shading) in inconsistent compared to consistent combined conditions (Fig. 3B), reliabilities did not differ between consistent and inconsistent conditions. Hence, the degree of stimulus inconsistency does not fully explain the observed lack of cue integration.

Zaidel et al. (2011) pointed out that in the case of systematic discrepancies, cues may undergo *mutual calibration*, whereby sensory cues adapt toward one another to achieve internal consistencies. The extent of adaptation that each cue undergoes has been proposed to either depend on its reliability (reliability-based cue adaptation (RBA)) or on its accuracy (fixed-ratio adaptation (FRA)). Our data suggest a shading-dominant adaptation with the motion cue undergoing large recalibration in combined cue conditions, i.e. the PSE for the motion cue shifting strongly towards that of the shading cue. Given that individual cues did not differ in terms of reliabilities, our results argue in favour of FRA as a possible mechanism contributing to shading-motion combinations.

An additional factor influencing cue-combined PSEs was accuracy of fixation. By comparing trials with small fixational eye movements (0-1.5 deg) to trials with moderate fixational eye movements (1.5-3 deg), we found that moderate eye movements led to a larger weighting of the motion cue (see Fig. 5). Successful suppression of fixational eye movements may have been due to the attentional suppression of the stimulus aspect triggering the eye movements, namely the moving dots. Hence, fixation might imply a reduction of motion processing and therefore an increased weight of the shading cue.

An interesting question pertains to the ecological factors underlying the differences in shading and motion cue weighting in our cue-combined data. It seems reasonable to state that fine judgments of 3D curvature in static objects are rarely based on motion or optic flow cues, presumably contributing to an experience-based reduction of the motion cue weight in combined cue stimuli. The role of experience in cue weighting is supported by a trend in our data towards more reliable motion estimates in the top compared to central and bottom fixation conditions that is consistent with previous findings of experience-based lower visual field biases (Previc 1990; Amenedo et al. 2007). These ideas are in line with Bayesian models of cue integration (van Ee et al. 2003) which can emulate cooperative cue integration, as predicted by MLE, but have the built-in flexibility to make cue integration conditional on satisfaction of constraints that reflect prior knowledge about the world (Battaglia et al., 2003; Knill, 2007; for a review see Vilares and Kording, 2011). One relevant constraint for cue integration in our study may be related to depth ambiguity. Bülthoff and Mallot (1987) have indicated that in case one cue provides ambiguous depth information, integration will only occur if the other cue can disambiguate that information. However, the cues studied here are known to be both ambiguous in terms of convexity/concavity (Orban 2011;

Lovell and Bloj 2012). Cooperative cue integration for the purpose of enhancing fine 3D curvature judgments studied here, may require disambiguation of the coarse object structure (e.g., convex/concave) and as suggested by our data, may be actively suppressed if this requirement is not fulfilled.

### **Neural correlates of and constraints for cue integration**

As exemplified by Ramachandran and Gregory (1987) for motion and colour, the presence (or lack) of cue interactions can be interpreted in the context of knowledge about the extent to which cues are segregated during early visual processing. Sincich and Horton (2005) have proposed a two-partite division of feedforward visual processing. One stream incorporates V1 patches, V2 thin stripes, and separate projection domains in V4, and is thought to be involved in the processing of wavelength/colour, as well as luminance/brightness, with a potential role in the processing of surface-related information (Hanazawa and Komatsu, 2001; Roe et al., 2005; for a review Goebel and De Weerd, 2009). The other stream, incorporating V1 interpatches, V2 pale and thick stripes, as well as separate projection domains in V4, may have a special role in processing of disparity and motion cues. Despite motion cues requiring longer processing time than disparity cues, both are processed within the same feedforward pathway, possibly providing a basis for their integration. Similarly, shading and disparity, although processed in separate processing streams, may integrate because both are processed quickly. Here however, shading and motion cues showed different processing times and additionally, according to Sincich and Horton (2005), are processed in separate feedforward streams. These factors together may reduce the likelihood of cue integration.

Furthermore, processing of shading-motion combinations in lower-order areas likely is modulated by feedback from higher-order areas that process both luminance/shading and motion 3D cues (Orban et al. 1999; Gerardin et al. 2010). This feedback may enable cooperative integration, or conversely may veto it when specific constraints are not met. Largely overlapping areas for 3D motion, disparity and texture cues have been demonstrated with fMRI in the IPS, ITG and LOS (Georgieva et al. 2008) and may potentially contribute to such feedback. Interestingly, the same study also showed that the processing of 3D shading was primarily restricted to the ITG, suggesting a special status for the extraction of shape-from-shading and additionally, a separation of shading and motion cues even at high levels of processing. An alternative route for feedback, particularly in our data, may involve IPS regions that have previously shown an involvement in signaling/detecting perceptual inconsistencies (Brouwer et al. 2009). From the IPS, information might have been redirected to lower cue-specific levels of the processing stream (Meinenbrock et al. 2007). In our data, the effect of such feedback signaling conflict might have been to veto integration and to remove conflict by decreasing the motion cue weight.

In sum, although MLE may be a broadly applicable mechanism for 3D cue integration, our results suggest that it does not hold for shading-motion cue combinations. Instead, our findings suggest that stimulus-related factors, perceptual

biases, attention, and prior world knowledge embedded in the visual system's structure and function may place constraints on statistically optimal cue integration schemes.

## References

- Amenedo E, Pazo-Alvarez P, Cadaveira F. 2007. Vertical asymmetries in pre-attentive detection of changes in motion direction. *Int J of Psychophys.* 64:184–189.
- Ban H, Preston TJ, Meeson A, Welchman AE. 2012. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci.* 15:636–643.
- Battaglia PW, Jacobs R a, Aslin RN. 2003. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am.* 20:1391–1397.
- Bradshaw MF, Rogers BJ. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res.* 36:3457–3468.
- Brouwer GJ, Tong F, Hagoort P, van Ee R. 2009. Perceptual incongruence influences bistability and cortical activation. *PLoS one.* 4:e5056.
- Bülthoff H, Mallot H. 1987b. Interaction of different modules in depth perception: Stereo and shading. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Technical report. 965.
- Clark J, Yuille A. 1990. Data fusion for sensory information processing systems. Boston: Kluwer Academic Publishers.
- Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature.* 415:429–433.
- Ernst MO, Bühlhoff HH. 2004. Merging the senses into a robust percept. *Trends Cogn Sci* 8:162–169. 8:162–169.
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. 2012. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci.* 15:146–154.
- Georgieva SS, Todd JT, Peeters R, Orban G a. 2008. The extraction of 3D shape from texture and shading in the human brain. *Cereb Cortex.* 18:2416–2438.
- Gerardin P. 2007. Shape from shading: New perspectives from the Polo Mint stimulus. *J Vis.* 7:1–11.
- Gerardin P, Kourtzi Z, Mamassian P. 2010. Prior knowledge of illumination for 3D perception in the human brain. *P Natl Acad Sci USA.* 107:16309–16314.
- Goebel R, Weerd PD. 2009. Perceptual filling-in: from experimental data to neural network modeling. In: Gazzaniga MS, editor. *The cognitive neurosciences IV.* MIT Press. p. 435–456.
- Gowen E, Abadi RV, Poliakoff E, Hansen PC, Miall RC. 2007. Modulation of saccadic intrusions by exogenous and endogenous attention. *Brain Res.* 1141:154–167.
- Hanazawa A, Komatsu H. 2001. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci.* 21:4490–4497.

- Hillis J, Watt S, Landy M, Banks M. 2004. Slant from texture and disparity cues: Optimal cue combination. *J Vis.* 4:967–992.
- Hillis JM, Ernst MO, Banks MS, Landy MS. 2002. Combining sensory information: mandatory fusion within, but not between, senses. *Science.* 298:1627–1630.
- Jacobs R a. 2002. What determines visual cue reliability? *Trends Cogn Sci.* 6:345–350.
- Jiang Y, Boehler CN, Nönnig N, Düzel E, Hopf J-M, Heinze H-J, Schoenfeld MA. 2008. Binding 3-D object perception in the human visual cortex. *J Cognitive Neurosci.* 20:553–562.
- Johnston EB, Cumming BG, Landy MS. 1994. Integration of stereopsis and motion shape cues. *Vision Res.* 34:2259–2275.
- Knill D. 2007. Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J Vis.* 7:1–24.
- Knill DC, Saunders J a. 2003. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis Res.* 43:2539–2558.
- Landy MS, Maloney LT, Johnston EB, Young M. 1995. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis Res.* 35:389–412.
- Langer M, Bulthoff H. 2001. A prior for global convexity in local shape-from-shading. *Perception.* 30:403–410.
- Lee YL, Saunders J. 2011. Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *J Vis.* 11:1–12.
- Lovell P, Bloj M. 2012. Optimal integration of shading and binocular disparity for depth perception. *J Vis.* 12:1–18.
- Meienbrock a, Naumer MJ, Doehrmann O, Singer W, Muckli L. 2007. Retinotopic effects during spatial audio-visual integration. *Neuropsychologia.* 45:531–539.
- Nachmias J. 1959. Two-Dimensional Motion of the Retinal Image during Monocular Fixation. *J Opt Soc Am.* 49:901–908.
- Orban G a. 2011. The Extraction of 3D Shape in the Visual System of Human and Nonhuman Primates. *Annu Rev Neurosci.* 34:361–388.
- Orban G a, Sunaert S, Todd JT, Van Hecke P, Marchal G. 1999. Human cortical regions involved in extracting depth from motion. *Neuron.* 24:929–940.
- Previc F. 1990. Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behav Brain Sci* 450:519–575.
- Ramachandran V, Gregory R. 1987. Does colour provide an input to human motion perception? *Nature.* 275:55–56.

- Regan D, Erkelens CJ, Collewijn H. 1986. Visual field defects for vergence eye movements and for stereomotion perception. *Invest Ophth Vis Sci.* 27:806–819.
- Rock I, Victor J. 1964. Vision and touch: An experimentally created conflict between the two senses. *Science.* 143:594–596.
- Roe AW, Lu HD, Hung CP. 2005. Cortical processing of a brightness illusion. *P Natl Acad Sci USA.* 102:3869–3874.
- Schiller PH, Slocum WM, Jao B, Weiner VS. 2011. The integration of disparity, shading and motion parallax cues for depth perception in humans and monkeys. *Brain Res.* 1377:67–77.
- Sincich LC, Horton JC. 2005. The circuitry of V1 and V2: integration of color, form, and motion. *Annu Rev Neurosci.* 28:303–326.
- Sun P, Freund R. 2004. Computation of minimum-volume covering ellipsoids. *Oper Res.* 52:690–706.
- Vilares I, Kording K. 2011. Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann NY Acad Sci.* 1224:22–39.
- Vuong QC, Domini F, Caudek C. 2006. Disparity and shading cues cooperate for surface interpolation. *Perception.* 35:145–155.
- Welchman AE, Deubelius A, Conrad V, Bülthoff HH, Kourtzi Z. 2005. 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci.* 8:820–827.
- Young MJ, Landy MS, Maloney LT. 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.* 33:2685–2696.
- Zaidel A, Turner AH, Angelaki DE. 2011. Multisensory calibration is independent of cue reliability. *J Neurosci.* 31:13949–13962.
- Zhang Y, Weiner VS, Slocum WM, Schiller PH. 2007. Depth from shading and disparity in humans and monkeys. *Visual Neurosci.* 24:207–215.
- van Ee R, Adams WJ, Mamassian P. 2003. Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J Opt Soc Am.* 20:1398–1406.



## CHAPTER 2

---

# An Exploration of the Neural Representations of Combined Shading and Motion Cues in Human Visual Cortex

---

*Corresponding publication:* Graewe B, Bastiani M, Jans B, Roebroek A, Castelo-Branco M, De Weerd P. An exploration of the neural representations of combined shading and motion cues in human visual cortex (*in preparation*).

## Abstract

To achieve robust estimates of depth, human observers typically exploit multiple cues. Here, we started to investigate the cortical mechanisms mediating 3D perception from combined shading and motion visual cues using a parametric curvature discrimination task and concurrent event-related functional magnetic resonance imaging (fMRI). Psychophysical data suggest that performance in both consistent combined (CC) and inconsistent combined (IC) stimulus conditions was a weighted average of performance in single cue conditions. This shows that, in contrast to many other depth cue combinations, combined shading and motion cues in the context of the present study did not integrate optimally. fMRI activity in group average data did not show consistent variations in response to stimulus curvature manipulations in neither the single nor the combined cue conditions. Although in individual participants, a dependency of fMRI activity on curvature could be established, this was difficult to relate to psychophysical performance. Interestingly, within selected ROIs in the right IPS, a combination of single cue fMRI-based curvature sensitivities was predictive of fMRI-based curvature-sensitivity in combined cue conditions. Although these predictions showed some weak correlations with performance measures, our findings are difficult to interpret at this stage. The lack of optimal integration of shading and motion for 3D perception in the performance data suggest an alternative competitive mechanism, a hypothesis we aim to test in future analysis using methods other than GLM-based approaches.

## Introduction

Sensory information offers multiple cues that can facilitate 3D perception and adept spatial interaction with the environment. To optimize depth estimates, cue combinations within the visual modality (e.g. texture and motion (Jacobs and Fine 1999), shading and disparity (Lovell and Bloj 2012) or motion and disparity (Johnston et al. 1994)) as well as across modalities (e.g. visual and haptic cues (Hillis et al. 2002)) have generally proven beneficial, resulting in increased sensitivity compared to single cue estimates. In many cases 3D cue combination data are best explained by a Maximum Likelihood Estimation (MLE) model (Clark and Yuille 1990), predicting that estimates based on combined cues are biased towards the more reliable single cue, and that the reliability of estimates based on the combined cue exceeds that of estimates based on single cues. Although neurophysiological studies in monkeys have shown convergence of information from 3D cues at lower (Bakin et al. 2000; Qiu and Heydt 2005) and higher levels (Liu et al. 2004) of the visual system, the neural basis of cue combination remains poorly understood. Human fMRI studies addressing cue combinations are scarce with the exceptions of Welchman et al. (2005) showing combined perspective and disparity cue representations in extrastriate areas MT and LOC, and additionally Ban et al. (2012) demonstrating combined depth signals from binocular disparity and relative motion in dorsal area V3B/KO.

While most psychophysical (Young et al. 1993; Hillis et al. 2004) as well as neurophysiological (Fetsch et al. 2012) and fMRI (Ban et al. 2012) evidence supports MLE as a mechanisms underlying 3D cue combinations, the combination of shading and motion cues has remained conspicuously understudied. Schiller et al. (2011) were the first to perform a psychophysical study investigating 3-way combinations of motion parallax, shading and disparity cues but the specific combination of shading and motion cues was not directly addressed. In a recent study (chapter 1), we provided the first evidence that shading-motion cue combinations, in contrast to other depth cue combinations, do not result in statistically optimal performance according to MLE. Instead, cue combination was better described by weighted averaging of estimates based on individual cues with no relationship between cue weights and individual cue reliabilities. Moreover, we found that the reliability of combined cue estimates was decreased compared to the reliability of single cue estimates. We speculated that stimulus-related factors as well as ecological and anatomical constraints may have precluded optimal cue combination. Instead, we observed a form of suppressive competition which we aimed to study here at the level of fMRI activation.

To explore mechanisms that may underlie shading-motion cue competition, we combined psychophysics and event-related fMRI to dissociate cue-specific cortical regions from those involved in coding combined shading-motion 3D percepts. Human observers performed a 3D surface curvature discrimination task based on single shading- and motion cues as well as based on combined cues that contained consistently or inconsistently combined curvature information. We

hypothesized that several dorsal-stream regions that have previously been implicated in parametric coding of stimulus depth from disparity (Preston et al. 2008), disparity and perspective (Brouwer et al. 2009) as well as from motion cues (Graewe et al. 2012), would reveal 3D curvature-dependent activation in response to single and combined shading-motion cues that would allow us to test cue combination/competition.

## Methods

### Participants

11 healthy volunteers (6 male; mean age 28.1 years,  $SD = 1.83$ ) with normal or corrected-to-normal vision participated in the study. Prior to the start of the experiment written and oral informed consent was obtained from all participants according to the declaration of Helsinki. Approval for this study was granted by the Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

### Stimuli

Stimuli for the cue combination experiment consisted of 3D ellipsoids (6 deg diameter) viewed with a gaze angle orthogonal to the flat surface that formed the background of the stimuli. Shapes were defined either by shading, motion or combined shading and motion cues and were shown on a light grey background (Fig. 1A). The x-y semi-principal axes of the ellipsoid formed the background plane, while the z-axis was orthogonal to the background and directed to the eyes of the observer. The x-y semi-principal axes always remained equal in length so that the shape's base appeared fully circular. The length of the z-axis was manipulated, and the resulting curvature changes were quantified on a normalized scale ranging from 0 to 10. A curvature of 10 describes a perfect half sphere protruding from the background by 3 deg towards the observer and a curvature of 0 describes a flat disk. In the experiment (Fig. 1A & 1C), the reference stimulus had 50% of the maximum curvature (curvature 5), whereas test stimuli had curvatures "2", "4", "5", "6" and "8". Curvature step 8 would correspond to 80% of the maximum curvature (see 'fMRI experimental design' for details on test and reference stimuli). Stimuli were created using MATLAB (R2009b, The MathWorks, Natick, Massachusetts, USA).

### *Shading-defined stimuli*

In shading stimuli (Fig. 1A, top row), the curvature of the ellipsoid was to be inferred from the luminance intensity distribution across the stimulus. To avoid participants using local cues to judge the curvature information, positions of the light sources were varied between consecutive stimuli within a trial (top left, top right, bottom left, bottom right). Thus, ellipsoids had one of 5 curvatures

(manipulation of the length of the z-axis) and were presented at one of the 4 light source positions, resulting in 20 different shading stimuli.

### ***Motion-defined stimuli***

The motion-defined ellipsoids (Fig. 1A, second row) were defined by a square motion flow field ( $8 \times 8$  deg) on a light grey background. Within this surface, a  $42 \times 42$  grid of possible dot starting points was defined, of which 90% was occupied, so that the flow field contained 1588 white dots ( $\sim 0.1$  deg diameter) moving coherently in an upward direction (from bottom to top of the flowfield). Dot lifetime was 250 ms (5 frames at play rate of 20 Hz), and dot birth was distributed over frames so that in each frame approximately 20% of the dots died and were randomly reborn. Each dot was born within a circular region centred on a grid point, with this circular region defined by a radius equal to 50% of the (closest) distance between neighbouring grid points. Spatial overlap of dots was avoided. Dots were created on the background plane and were then projected onto the ellipsoid so that in the frontal perspective of the observer there was a fixed spacing between adjacent grid points. Moreover, dot size and shape was constant. Hence, dot density, size and shape did not provide a cue to depth perception. By contrast, each individual dot moved at a constant speed on the surface of the shape (4 deg/s), leading to differences between dots' speed in the observer's frontal view. Thus, gradients in dot speed in the observer's frontal view were the only cue from which surface curvature could be derived. As in the shading-defined stimuli, ellipsoids had one of 5 curvatures.

### ***Cue combined stimuli***

In combined cue stimuli, a motion-defined ellipsoidal stimulus was superimposed on a shading defined stimulus of equal size. As in the single shading stimuli, light source position of the shading cue was varied to avoid the use of local stimulus cues for curvature judgments. There were 20 (5 curvature steps  $\times$  4 light sources) 'consistent' combined stimuli in which shading and motion described the same curvature (e.g. shading 4 paired with motion 4) (Fig. 1A, third row) and 16 inconsistent stimulus combinations in which shading and motion cues described incongruent curvatures (Fig. 1A, fourth row), i.e. "shading 2; motion 8", "shading 4; motion 6", "shading 6; motion 4", "shading 8; motion 2" (see Fig. 1C)  $\times$  4 light source positions.

### **Design and Procedure**

Stimuli were delivered and synchronized with MR pulses using the software package Presentation (Neurobehavioral systems). Images were projected by an LCD projector (Sanyo PLC-ET30L) onto a screen (Samsung Syncmaster 931BF, refresh rate 60 Hz) at the head of the participants and were viewed via an angled mirror mounted to the head coil at a viewing distance of 57 cm. Responses were given via an MR compatible button box. Each volunteer participated in two scanning sessions (1 localizer session and 1 cue combination experimental session) on separate days.

## **Functional Localizers**

### *Retinotopic mapping*

Borders of the visual areas were defined by polar angle mapping according to standard techniques (Serenio et al. 1995). Participants viewed a single rotating red/green checkerboard wedge while fixating at a fixation cross at the centre of the screen. To facilitate accurate fixation, the stimulation was divided into 4 runs of 140 volumes.

### *Motion localizer*

To localize motion sensitive area V5/hMT+, we used a standard procedure with two main conditions showing either 400 dots moving radially outwards from the centre of the screen or showing 400 stationary dots (Goebel et al. 1998). Participants viewed four blocks (duration 8 s) of each experimental condition and were instructed to watch attentively while keeping central fixation. Each alternate block contained a central fixation cross and served as a rest condition.

### *Parietal network localizer*

For the localization of attention-related depth-sensitive parietal areas, we used alternating blocks of central fixation and saccadic eye movements. In saccadic eye movement blocks, participants were instructed to saccade to a white circle that could appear at one out of eight pre-defined locations along the horizontal or vertical meridian (alternation rate 500 ms) on a black background. There were 21 blocks lasting 16 seconds each.

### *Object Localizer*

Localizer stimuli for defining LOC and PPA were static grayscale images of faces, places (landscapes, skylines), objects (tools, cars, chairs), and scrambled versions of objects ( $9.5 \times 9.5$  deg). Stimuli were presented in the middle of the screen on a dark background and contained a central fixation cross (radius  $0.15^\circ$ ). There were 20 s long stimulation blocks (20 images, shown for 800 ms, ISI of 200 ms) separated by fixation blocks of 10 s. Each category block was repeated 4 times resulting in a single 8.2 min run. To maintain stable attention levels participants performed a one-back task.

## ***fMRI experimental design and task***

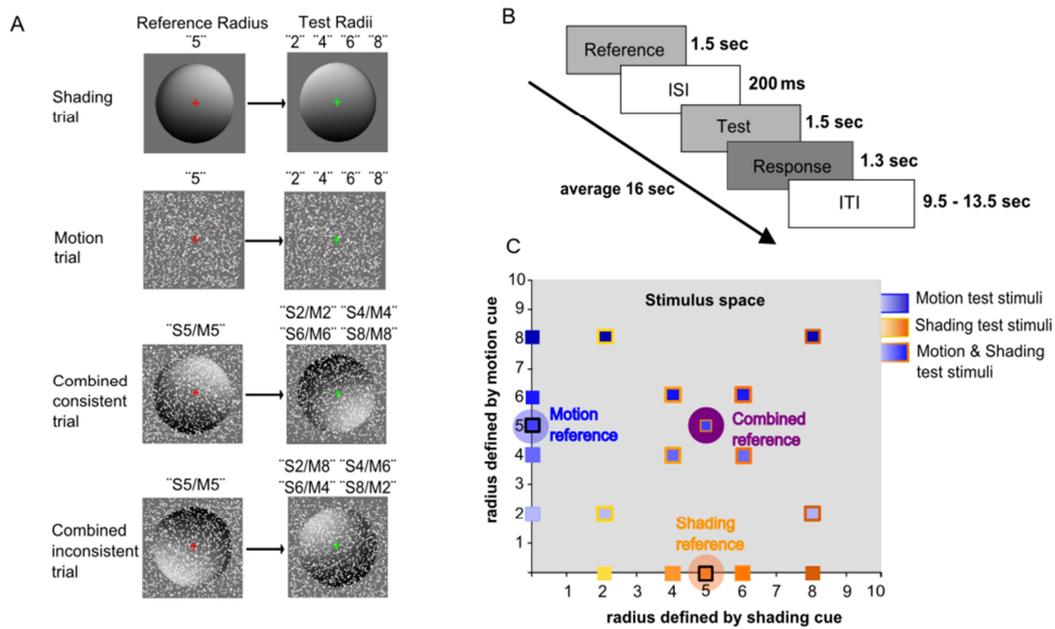
We used a slow event-related experimental design consisting of 4 sequential runs of 17 min containing 4 trials per condition (16 trials per condition in total). Each trial lasted 4.5 s (Fig. 1B), starting with the presentation of a reference stimulus (1.5 s) followed by an inter-stimulus-interval (ISI) (200 ms) and a test stimulus (1.5 s). Participants were asked to compare the curvature of the reference shape to the curvature of the test shape and indicated in a subsequent response period (1.3 sec) whether the test shape conveyed “more curvature” (right middle finger) or “less curvature” (right index finger). No feedback was given. To avoid misinterpretation

of the status of the presented stimuli, reference and test stimuli were indicated to the observers by red and green fixation crosses, respectively (Fig. 1A). During the ISI the fixation cross remained red until the test stimulus appeared. Trials were separated by a variable inter-trial-interval (ITI) of 9.5, 11.5 or 13.5 s.

There were 4 different trial types that were randomly presented throughout a run (Fig. 1C): In single shading (S) and single motion (M) trials the curvature in reference and test stimuli was defined by the same single cue (i.e. either shading or motion). In consistent combined (CC) trials the curvature in reference and test stimuli was defined by consistent combined cues. In inconsistent combined (IC) trials the curvature of the reference was defined by consistent combined cues whereas the curvature of the test stimuli was defined by inconsistent combined cues. Reference stimuli in all trial types conveyed a curvature of “5” (defined either by single cues or by CC cues). Test stimuli in single and CC cue trials could convey curvatures “2”, “4”, “6” and “8” and in inconsistent combined trials they could convey curvatures “shading 2; motion 8”, “shading 4; motion 6”, “shading 6; motion 4”, “shading 8; motion 2”. Prior to the start of the fMRI experiment, participants completed a practice run outside of the scanner to ensure comprehension of stimuli and task. Figure 1C summarizes the stimulus combinations and trial structure.

### ***fMRI acquisition***

Brain imaging was performed using a 3 Tesla Siemens Allegra MR head scanner at the Maastricht Brain Imaging Center. Participants lay supine in the scanner with their head stabilized by soft foam pads to reduce involuntary head motion. The functional localizer session consisted of one MT localizer run (102 volumes), one parietal network localizer run (169 volumes) and one object localizer run (248 volumes) with the following scanning parameters: EPI, TR = 2 s, TE = 30 ms, interleaved slice acquisition order, voxel size 3.5×3.5×3.5, 32 slices, in-plane matrix 64×64, Flip angle 90 deg, FOV=224 mm. Additionally we performed polar angle mapping to define retinotopic areas (4 sequential blocks of 140 volumes, EPI, TR = 2 s, TE = 30 ms, interleaved slice acquisition order, voxel size 2×2×2, 28 slices, in-plane matrix 112×112, Flip angle 90 deg, FOV=224 mm). The cue combination session consisted of 4 sequential runs of the cue combination experiment (length of each run: 522 volumes; parameters: EPI, TR = 2 s, TE = 30 ms, interleaved slice acquisition order, voxel size 3.5×3.5×3.5, 29 slices, in-plane matrix 64×64, Flip angle 90 deg, FOV = 224 mm). In both sessions we also acquired a T1-weighted anatomical scan (MPRAGE, TR = 2.25 s, TE = 2.6 ms, 192 slices, voxel size 1×1×1, Flip angle 9 deg, FOV = 256 mm).



**Fig. 1 Paradigm and conditions.** **A** Stimulus conditions. Throughout the cue combination experiment, there were 4 trial types (2 single cue and 2 combined cue trials), which were presented randomly. In each trial, a reference stimulus (curvature 5), was followed by a test stimulus. Reference and test stimuli in the experiment were indicated by red and green fixation crosses, respectively. **B** Single-trial timing. Reference and test stimuli were presented for 1.5 s each, separated by a 200 ms inter-stimulus-interval (ISI). During the response period participants indicated by button press whether the curvature of the test stimulus was smaller or larger than the curvature of the reference stimulus. Trials were separated by an inter-trial-interval (ITI) of variable duration. **C** Stimulus space showing the reference and test stimuli in each of the 4 conditions. Single cues are color-coded in orange (shading, x-axis) and blue (motion, y-axis) with the range from lighter to darker colors representing an increase in the shapes' curvatures. Consistent and inconsistent combined conditions are represented by combined blue and yellow/orange symbols on the two diagonal axes. Reference stimuli are indicated by orange, blue or purple circles.

## Analysis

### **Analysis of Psychophysical data**

#### *Accuracy*

For each cue type, the proportion of “larger curvature responses” was plotted as a function of stimulus curvature increase in each participant. Linear trend lines best fitting the data points in each condition were then computed and their slopes were used as a measure of performance-based curvature sensitivity ( $cs$ ) in the respective cue condition.  $cs$  differences between conditions were established by repeated-measures ANOVA with the factor cue type (S, M, CC, IC) and in the case of significant main effects were further specified by computing pairwise comparisons. Additionally, we estimated cue weights for shading and motion cues based on their performance-based  $cs$  (slopes) in the combined conditions. Following a linear model of cue combination we postulated that  $cs$  in the combined condition should be a weighted sum of single motion and shading slopes ( $cs$ ), where the sum of the weights add up to 1 ( $w_m + w_s = 1$ )

$$Slope_{cc} = w_m Slope_m + w_s Slope_s \quad (1)$$

Hence, the motion weight to the combined consistent slope is given by

$$W_m = \frac{Slope_{cc} - Slope_s}{Slope_m - Slope_s} \quad (2)$$

whereas the motion cue weight to the inconsistent combined slope is calculated by

$$W_m = \frac{Slope_{ic} - Slope_s}{-Slope_m - Slope_s} \quad (3)$$

Recall that in the inconsistent combined cue pairs the motion described curvature range was flipped compared to the shading cue (see Fig. 1C). Hence the slope of the motion cue in the IC condition is given by the negative slope of the single motion cue.

#### *Response Times*

Trials with response times that were  $\pm 2$  SD from the mean of the respective condition were labelled as outliers and were removed from the psychophysical response time (RT) analysis. Average RT data for each condition and participant were subjected to a repeated-measures ANOVA with factors curvature (1-4) and cue type (S, M, CC, IC). RT-based cue weights in the CC and the IC conditions were calculated by pooling over all curvature conditions. Motion cue weights in the CC condition were derived based on single cue RTs and combined cue RTs ( $w_m \times RT_M + (1-w_m) \times RT_S = RT_{CC}$ ). Solving this equation for  $w_m$  yielded the motion cue

weights in the CC condition. Recall that  $w_s$  was given by  $1-w_m$ . The same calculations were performed to derive single cue RT weights in the IC condition ( $w_m \times RT M + (1-w_m) \times RT S = RT IC$ ).

When applicable, p-values were corrected for non-sphericity using the Greenhouse-Geisser correction. Statistical post-hoc tests were performed with a Bonferroni multiple-comparisons correction.

### ***Analysis of fMRI data: Pre-processing and ROI definition***

#### *Pre-processing*

fMRI data were analyzed using BrainVoyager QX (Brain Innovation). Pre-processing of the functional data consisted of slice-scan time correction, head motion correction (using the 'trilinear/ sinc interpolation' option), temporal high-pass filtering (3 cycles) and linear trend removal. The first 2 volumes were discarded from the analysis to allow for scanner equilibration. Spatial smoothing (4 mm) was only applied in the group data analysis. Individual functional runs were co-registered with the participants' corresponding anatomical scans and were transformed into Talairach space. Individual cortical surface reconstructions were derived from segmentations at the gray/white matter boundary using a semi-automatic procedure based on intensity values. For whole-brain group analysis anatomical and functional data were aligned using the cortex-based intersubject alignment (CBA) procedure (moving-target group average approach). This method improves the anatomical alignment between gyri and sulci across subjects and increases spatial specificity of group analyses (e.g. van Attevelde, Formisano, Goebel, & Blomert, 2004). Multiple comparisons were corrected for using cluster-size thresholding (Forman et al. 1995; Goebel et al. 2006). The minimum cluster-size threshold that yielded a cluster-level false positive rate ( $\alpha = .05$ ) after 500 iterations was applied to the statistical contrast maps of the functional data.

#### *ROI Definition*

The lateral occipital complex (LOC) was defined by the contrast objects > matched scrambled objects and the parahippocampal place area (PPA) was defined by the contrast scenes > faces and objects. Medial temporal area MT was identified using the contrast moving dots > stationary dots. Mean Talairach coordinates were in agreement with previous studies (rLOC:  $N = 10$ ,  $x = 42$ ,  $y = -77$ ,  $z = -11$ ; lLOC:  $N = 9$ ,  $x = -45$ ,  $y = -76$ ,  $z = -7$ ; rPPA:  $N = 11$ ,  $x = 22$ ,  $y = -40$ ,  $z = -12$ ; lPPA:  $N = 11$ ,  $x = -20$ ,  $y = -42$ ,  $z = -11$ ; rMT:  $N = 11$ ,  $x = 44$ ,  $y = -66$ ,  $z = -2$ ; lMT:  $N = 11$ ,  $x = -45$ ,  $y = -66$ ,  $z = 1$ ). In addition to ventral object related regions and motion sensitive area MT, we also aimed to study dorsal stream areas that have been shown to be involved in high level 3D shape perception (Gerardin et al. 2010). Contrasting blocks of fixation with blocks of saccadic eye movements showed several areas in the intraparietal sulcus (IPS), i.e. ventral intraparietal sulcus (VIPS), dorsal intraparietal sulcus (DIPS) and parieto-occipital intraparietal sulcus (POIPS). However, since POIPS could only be localized in 6 (of 11) participants, analysis was restricted to VIPS and DIPS. Polar

angle mapping enabled us to delineate the borders between areas V1, V2, V3, V3A, VP and V4. Additionally, area V3B/KO which is located posterior to MT, was defined retinotopically as the region of cortex located inferior to and sharing a border with V3A (Dupont et al. 1997; Preston et al. 2008).

### ***Analysis of fMRI data: ROI statistical analysis***

For the statistical analyses in our ROIs, data from retinotopic regions were pooled across left and right hemispheres, whereas separate beta values for left and right hemispheres were analysed in MT, PPA, LOC and the two IPS regions (VIPS and DIPS). This was motivated by previous research indicating hemispheric differences in regions that are implicated in high-level object perception (Dien 2009), while, in the scope of the present task, we did not predict such differences in retinotopic regions.

#### *Overall effect of cue combinations*

In a first ROI analysis, we aimed to establish the overall effect of combining cues in consistent and inconsistent ways. To this end, we pooled across curvature levels and performed separate GLMs for each ROI testing for a main effect of cue type (shading, motion, CC, IC). In retinotopically-defined regions, beta values were averaged across left and right hemispheres, whereas in MT, LOC, PPA and IPS regions, hemisphere was included as an additional factor in the analysis.

#### *Establishing curvature sensitivities in individual ROIs*

In each ROI we tested for a main effect of curvature from the 4 cue types using repeated-measures ANOVA (based on averaged hemispheric data in retinotopic regions and in separate analyses per hemisphere in the remaining ROIs). Additionally for subsequent analyses, fMRI-based curvature sensitivity in each of the 4 cue types was quantified in single participants by fitting linear trend lines to beta values plotted as a function of curvature. The slopes of these lines were used as a measure of fMRI-based curvature sensitivity (*cs*). Note that these slopes could be *positive* in case the beta values *increased* as a function of curvature and *negative* in case the beta values *decreased* as a function of curvature. Therefore the concept of negative *cs* refers to stimulus curvature increases that are coded by a decrease in beta values.

#### *Multiple linear regression analysis in individual ROIs*

ROI analysis revealed that beta values in different participants could increase or decrease as a function of curvature. Moreover, whether in the same participant beta values increased or decreased as a function of curvature also depended on the cue type that defined the curvature. Although this would suggest a null-relation between beta values and curvature in average data, individual data did suggest a sensitivity of the fMRI activation to curvature. Therefore, we intended to examine whether individual participants' fMRI-based *cs* in single cue conditions predicted fMRI-based *cs* in combined cue conditions. To that goal, we conducted multiple

linear regression analyses. First, we sought to test whether the fMRI-based *cs* in the CC condition (dependent variable) could be explained by a combination of single shading and motion cue fMRI-based *cs* (predictors). Secondly, we performed the same analysis using fMRI-based *cs* in single shading and single motion cues as predictors but using fMRI-based curvature sensitivity in IC conditions as dependent variable. These two analyses were performed in all localized ROIs. At the initial stage of analysis the regression model always contained single cue *cs* as well as an interaction term which was however discarded in case it did not contribute to explaining the variability in the dependent variable.

### *Correlations with psychophysical data*

In a follow-up analysis we tested how performance data and RTs were related to fMRI-based *cs*. Thereto we restricted our analysis to regions that in the previous analysis showed to combine single cue curvatures, i.e. rPPA, rVIPS and rDIPS.

First, we correlated performance-based *cs* with fMRI-based *cs* of the same cue type in each ROI. Recall that performance-based *cs* corresponded to the slopes of linear trend lines fitted to proportions of 'larger' responses as a function of curvature. Similarly, fMRI-based *cs* corresponded to the slopes of linear trend lines fitted to beta values as a function of curvature.

Secondly, since fMRI-based *cs* in the combined cue conditions was best explained by a combination of single cue fMRI-based *cs*, we here aimed to also establish a relationship between single cue fMRI-based *cs* and performance in the combined cue conditions. To that goal, we used multiple linear regression analyses with fMRI-based *cs* in the single shading and motion conditions as predictors and performance-based *cs* in the combined condition as dependent variable. Separate analyses were conducted to predict performance in the CC and IC conditions.

As an additional measure, we computed *cs* cue weights in each ROI by first computing the *absolute* difference between fMRI-based shading and motion *cs* (slopes). Then, we gave a positive sign to this difference value to indicate shading dominance (greater absolute slope for shading cue than for motion cue, i.e. irrespective of the signs of the slopes). Likewise, we gave a negative sign to this difference value to indicate motion dominance (greater absolute slope for motion cue than for shading cue). The value of the resulting *cs* cue weight represents the absolute difference in slope between the two cues and its sign indicates whether the slope was steeper for shading (positive) or for motion (negative). The resulting *cs* cue weights were then correlated with the performance-based *cs* in the CC and IC conditions.

Finally, we also aimed to establish a correlation between fMRI-based *cs* and RTs. Thus, we first correlated RTs with fMRI-based *cs* of the same cue type and additionally, we performed multiple linear regression using single cue fMRI-based *cs* as predictors and RTs in the combined cue conditions as dependent variable. Again, separate analyses were conducted to predict RTs in the CC and IC conditions.

### **Analysis of fMRI data: Whole brain statistical analysis**

We also sought to trace regions specifically involved in representing combined shading-motion representations outside of our independently localized ROIs. To that goal, we performed a random-effects (RFX) whole brain analysis using several conjunction contrasts. First, we identified regions that responded more strongly to consistently combined stimulation than to each single cue condition (pooled over all curvature levels) by performing the contrast (combined consistent > single shading)  $\cap$  (combined consistent > single motion) (C1). We further constrained this contrast by excluding regions that were not activated by each single cue compared to baseline (C1  $\cap$  (single shading > baseline)  $\cap$  (single motion > baseline)) (C2). Additionally, to detect regions that are sensitive to curvature inconsistencies, we contrasted all consistent conditions with all inconsistent conditions (C3).

## **Results**

### **Psychophysics**

#### **Accuracy**

Psychophysical performance in each cue condition was assessed by plotting the proportion 'larger' responses (test stimulus judged larger than the reference) as a function of curvature (Fig. 2A). Subsequently we fitted linear trend lines to the data and entered the corresponding slopes (performance-based  $c_s$ ) of the 4 conditions into a repeated-measures ANOVA. The significant main effect of cue type ( $F_{(3,30)} = 13.672, p = .001$ ) corresponded to better performance in the single shading and CC conditions compared to the single motion condition ( $p = .01$  and  $p = .036$  respectively). Moreover, slopes were steeper in shading compared to IC conditions and steeper in CC compared to IC conditions ( $p = .013$ ). The remaining comparisons did not prove significant ( $p > .05$ ).

Hence,  $c_s$  (steepness of the slope of linear trend lines) was greater in single shading compared to single motion cues. Performance in the CC condition differed significantly from single motion but not from single shading performance, suggesting a dominance of the shading cue to the CC percept. This fits with performance in the IC condition, in which the positive sign of the regression line slope in the group average (slope  $\beta = 0.1$ ) suggests a stronger weighting of the shading compared to the motion cue. However, the decrease in the slope of the regression line in IC compared to CC trials suggests that in the IC condition a weighted combination of curvatures defined by both single cues were perceived.

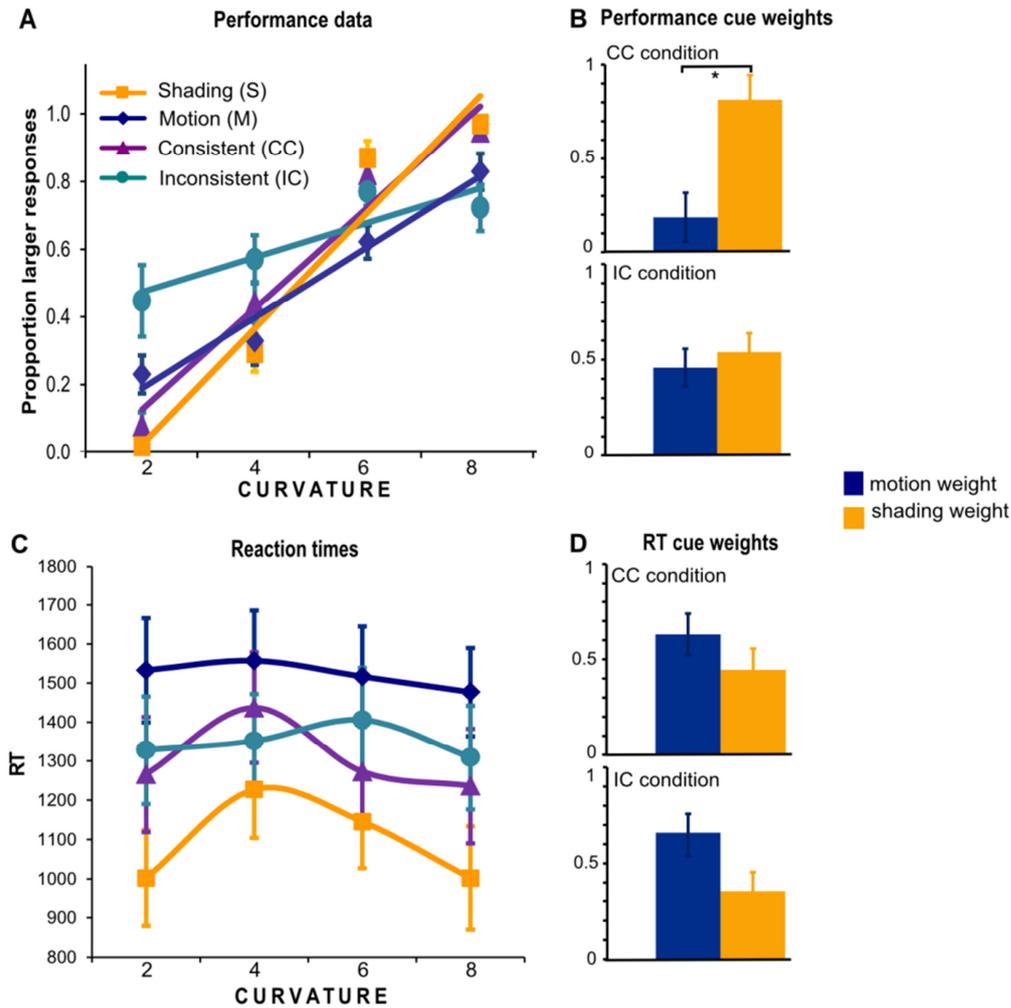
In agreement with IC and CC slopes, cue weight estimation (see Methods) based on the CC condition confirmed a significant dominance of the shading cue ( $w_s = 0.81$ ) over the motion cue (t-test on cue weights:  $t_{(10)} = 2.403, p = .037$ ) whereas in the IC condition cues were weighted approximately equally with only a slight dominance of the shading cue ( $w_s = .54$ ) (t-test on cue weights:  $p > .05$ ) (Fig. 1B). While this indicates a trend towards relatively larger cue weighting of the motion cue in the

inconsistent compared to consistent condition, differences between cue weights in CC and IC condition missed statistical significance ( $p > .05$ ).

### **Reaction Times**

To test RT differences between conditions, a repeated-measures ANOVA with the factors cue type (S, M, CC, IC) and curvature (4 levels) was computed, revealing significant main effects of both factors ( $F_{(3,30)} = 26.031, p < .001$  and  $F_{(3,30)} = 6.334, p = .007$ , respectively) as well as a significant interaction term ( $F_{(9,90)} = 4.147, p < .001$ ). This interaction pertained to an effect of curvature only in the shading and consistent combined conditions ( $F_{(3,30)} = 7.597, p = .007$  and  $F_{(3,30)} = 9.296, p < .001$ , respectively), showing slower RTs for “difficult” (curvature levels “4” and “6”) compared to easy conditions (curvature levels “2” and “8”). This task difficulty effect on RT was due to conditions “4” and “6” strongly resembling the curvature level of the reference (“5”). Furthermore, RTs were longest in single motion conditions, intermediate in combined conditions (CC and IC) and fastest in single shading conditions (Fig. 2C).

Cue weight estimation based on RTs showed a similar pattern for both CC and IC conditions with a slightly larger influence of the motion compared to the shading cue ( $w_s = .44$  and  $w_s = .34$ , respectively) (Fig. 2D). However, the differences between shading and motion cue weighting in RT data were not statistically significant ( $p > .05$ , in CC and IC conditions). Similarly, cue weighting did not differ significantly between CC and IC condition ( $p > .05$ ).



**Fig. 2 Performance-based curvature sensitivity.** **A** Performance data. For each condition, the proportion of responses in which the test stimulus was perceived larger than the reference was plotted as a function of curvature. Slopes of the best fitting linear trend lines were used as a measure of participants' performance-based *cs*. **B** Summary of performance cue weights. Average shading and motion cue weights were determined in CC and IC conditions based on single participant's performance-based *cs* (slopes). **C** Reaction time data. RTs were longest in single motion and shortest in single shading conditions while both combined conditions (CC and IC) resulted in intermediate RTs. In single shading and CC conditions RTs also tended to be longer for more difficult curvature conditions (curvature '4' and '6'). **D** Summary of RT cue weights. Shading and motion cue weights based on RTs in CC and IC conditions suggested a slight dominance of the motion cue but did not differ significantly ( $p > .05$ ). Error bars represent the standard error of the mean (SEM).

## ROI Analysis

### **Overall effects of cue combination**

#### *Areas defined by retinotopic mapping*

Since no differences between left and right hemispheres were expected in retinotopically-defined regions, analyses were performed on averaged data from both hemispheres. Results revealed a significant main effect of cue type (all  $p$ 's < .001) in all ROIs (V1, V2, V3, V3A, V3B, VP, V4) (Fig. 3A). Pairwise comparisons (after

corrections for multiple comparisons) revealed in all ROIs significantly smaller activation for the shading cue compared to motion and combined cues (all  $p$ 's < .05). In area VP, additionally, the CC condition showed significantly larger activation than the IC condition ( $p = .038$ ) (Fig. 3A).

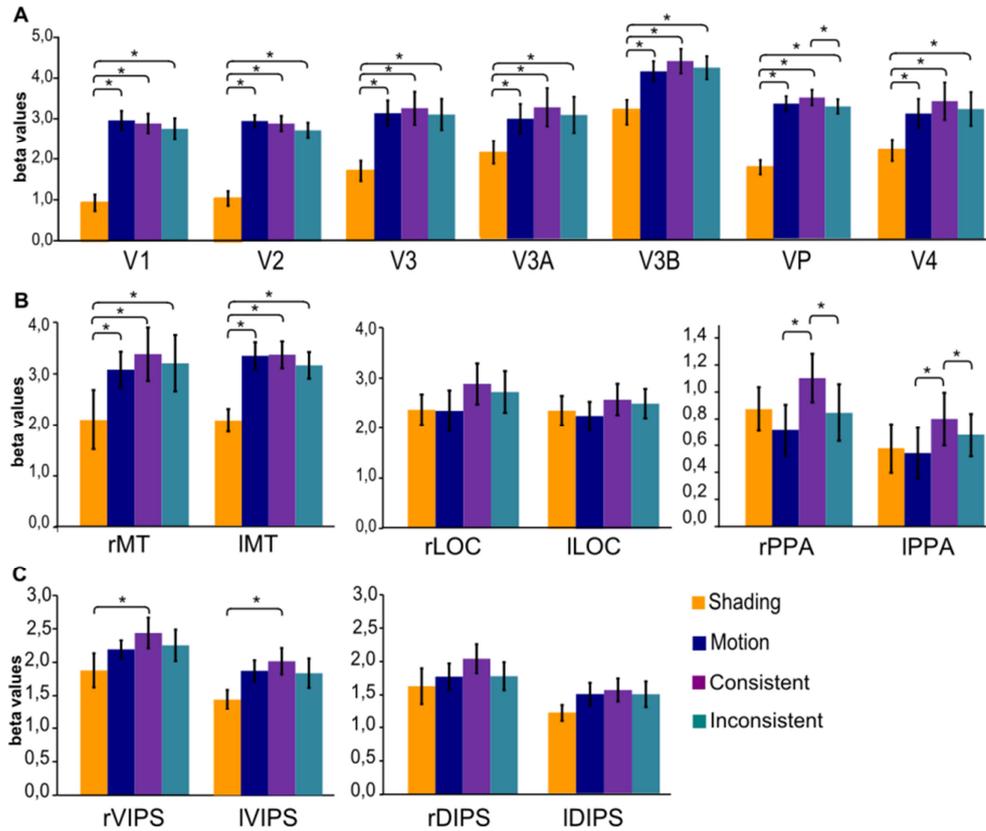
#### *Areas defined by object or motion localizers*

Since hemispheric differences have previously been reported for regions that are implicated in high-level object perception (Dien 2009), ANOVAs in MT, LOC, PPA and IPS included the factor cue type and additionally the factor hemisphere. Results showed a significant main effect of cue type in MT ( $F_{(3,30)} = 33.317$ ,  $p < .001$ ), corresponding to significant pairwise comparisons between shading and all other cue conditions (all  $p$ 's < .001) (Fig. 3B, left panel). Hemisphere (left, right) did not show a significant main effect or interaction with cue type ( $p > .05$ ). Hence, MT was more activated by stimuli containing motion signals compared to shading but did neither distinguish between combined cue and single motion conditions, nor between CC and IC stimuli. Interestingly, area LOC did not show main effects of cue type or hemisphere, nor interactions (all  $p$ 's > .05) (Fig. 3B, middle panel). Area PPA revealed a significant main effect of cue type ( $F_{(3,30)} = 5.921$ ,  $p = .003$ ) but no effect of hemisphere or interactions ( $p > .05$ ). In particular, the activation proved significantly greater in CC conditions than in single motion ( $p = .001$ ) and IC conditions ( $p = .030$ ) (Fig. 3B, right panel).

#### *Parietal network: VIPS & DIPS*

The VIPS showed both a significant main effect of cue type ( $F_{(3,30)} = 8.104$ ,  $p < .001$ ) as well as of hemisphere ( $F_{(3,30)} = 5.606$ ,  $p = .039$ ) (Fig. 3C, left panel). Shading cue-related activation was smaller than CC cue-related activation ( $p = .014$ ) while the remaining pairwise comparisons did not prove significant ( $p > .05$ ). The effect of hemisphere was due to the right hemisphere showing overall larger activation than the left hemisphere. In the DIPS main effects or interactions between conditions were not statistically reliable ( $p > .05$ ) (Fig. 3C, right panel).

In sum, retinotopic areas including MT displayed larger activation for stimuli containing motion but were unaffected by cue combinations. One exception to this result was area VP which showed more activation in CC than in IC conditions. In addition, PPA was also activated more strongly in CC compared to IC conditions (and also more activated in CC than in single motion conditions), suggesting a specific sensitivity in PPA and VP to agreement (or conversely conflict) between information given by shading and motion cues. Overall, beta values in midlevel retinotopic areas V3A, V3B, VP, V4 as well as MT, LOC and parietal areas showed a small trend towards larger activation to CC compared to both single cues (see Fig. 3A-C), which however did not survive correction for multiple comparisons.



**Fig. 3 Results from ROI analyses** in **A** retinotopically-defined regions, **B** bilateral motion and object-selective regions and **C** bilateral ventral and dorsal regions in the IPS. Bar graphs indicate fMRI activity in each of the 4 stimulus conditions pooled over curvature levels. Brackets and asterisks denote significant pairwise comparisons ( $p < .05$ ). Error bars indicate the SEM. Abbreviations refer to middle temporal area (MT); the lateral occipital complex (LOC); the parahippocampal place area (PPA); the ventral intraparietal sulcus (VIPS) and the dorsal intraparietal sulcus (DIPS).

### **Coding parametric changes in stimulus curvature**

In a subsequent analysis, we sought to establish each ROI's sensitivity to parametric changes in stimulus curvature conveyed by single and combined cues. Separate GLMs in each ROI were computed testing for a main effect of curvature from each cue (4 levels). As in the above reported analyses, in retinotopically-defined regions (V1, V2, V3, V3A, V3B, VP, V4), analyses were performed on averaged hemisphere data. For the remaining regions (MT, LOC, PPA and IPS), separate analyses for left and right hemispheric ROIs were computed. In accordance with results of previous studies (Brouwer et al. 2009; Gerardin et al. 2010) no significant curvature sensitivity in either of the cues was found in early retinotopically-defined areas (all ROIs  $p > .05$ ). In addition, none of the group GLMs in bilateral MT, LOC, PPA or parietal regions (VIPS and DIPS) revealed significant main effects of curvature ( $p > .05$ ) (see Fig. 4A & B for average group activation as a function of curvature in rVIPS and rDIPS).

However, single participant data in MT, PPA and IPS did show dependencies of beta values on curvature, for shading, motion and combined cues (insets in Fig. 4C & D). Surprisingly, the direction of this relationship (beta values increasing or decreasing as a function of curvature) varied in a non-systematic manner across cues and participants (Fig. 4C & D), thus yielding a zero-effect in data averaged across participants. To quantify the direction and strength of the relationship between beta values and curvature in single participants, we fitted, for each cue type, linear trend lines to beta values plotted as a function of the 4 curvature conditions. The sign and the slope of the lines represented fMRI-based  $cs$  in a respective condition, i.e. fMRI-based  $cs$  would be positive in case beta values increased as a function of curvature, and negative in case beta values decreased as a function of curvature. Main plots in Figure 4C and 4D display each of the 11 participants' fMRI-based  $cs$  to single and consistent combined cues in rVIPS and to single and inconsistent combined cues in rDIPS, respectively.

For example, in rVIPS (Fig. 4C), participant AL (inset left side), displayed increasing beta values as a function of motion-defined curvature but decreasing beta values as a function of shading-defined curvatures. Hence, the line fitted through the beta values of the single motion condition (blue) had a positive slope ( $\sim 0.2$ ), resulting in a 0.2-value for the blue symbol in the main plot below. Similarly, the line fitted through the beta values of the single shading condition (yellow) had a negative slope ( $\sim -0.4$ ), resulting in a -0.4 value for the yellow symbol in the main plot below. Finally, the same procedure was followed to generate a slope value for the CC condition (purple). Note that participant VK (Fig. 4C, inset right side), showed an increase in beta values as a function of curvature in all cue conditions, resulting in positive slopes, indicated accordingly for the single cue and CC conditions in the rVIPS main plot. To organize the data of the population of participants, we ordered data points according to  $cs$  in the shading condition, starting from most negative values on the left, to most positive values on the right. For illustrative purposes, we also connected data points belonging to the same condition across individuals. In Figure 4D, a similar graph was generated for rDIPS, this time only showing results for the two single and the IC cue conditions, in a format identical to that in Figure 4C.

Taken together this analysis shows that in the group average there are no systematic changes in the fMRI signal as a function of curvature. In individual participants, there were such relationships between fMRI signal and curvature, but the sign of this relationship could be negative or positive in both single and combined cue conditions.

### ***Multiple linear regression analysis in individual ROIs***

We hypothesized that fMRI activity in a region that is sensitive to curvature from each single depth cue would display a combination of single cue  $cs$  in combined cue conditions. To test for a systematic relationship between participants' single cue fMRI-based  $cs$  and combined cue fMRI-based  $cs$  (slopes), we performed a multiple linear regression analysis with shading and motion fMRI-based  $cs$  as predictors and

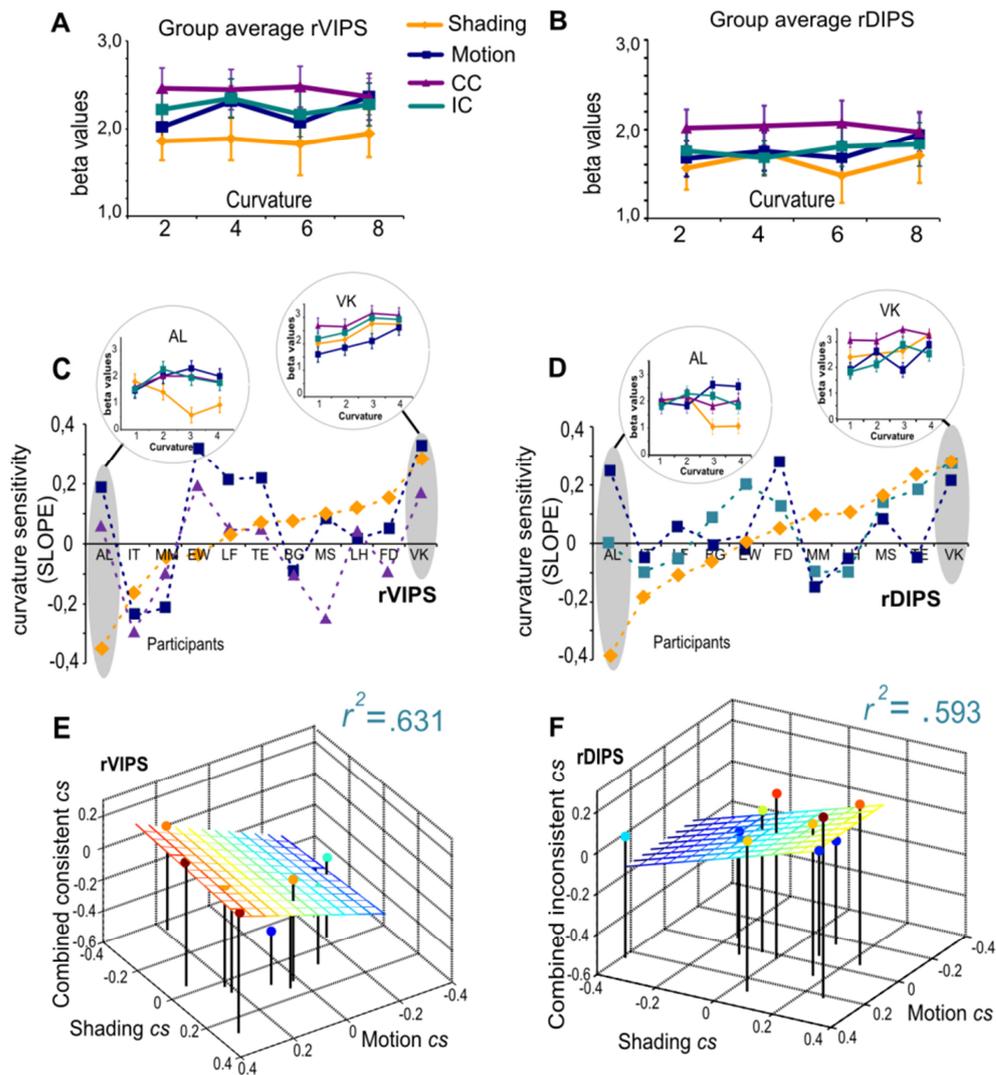
the combined consistent fMRI-based *cs* as dependent variable. Interestingly, we found that in the rVIPS, the fMRI-based *cs* of the CC condition could be predicted based on the fMRI-based *cs* of the single cues ( $r^2 = .631$ ,  $\beta_{\text{shading}} = -.024$ ,  $\beta_{\text{motion}} = .635$ ,  $F_{(2,10)} = 6.827$ ,  $p = .019$ ) (Fig. 4E). A similar pattern albeit less significant was also observed in rPPA ( $r^2 = .533$ ,  $\beta_{\text{shading}} = .119$ ,  $\beta_{\text{motion}} = .770$ ,  $F_{(2,10)} = 4.568$ ,  $p = .048$ ) (not shown). Analysis in the remaining ROIs pertaining to prediction of fMRI-based *cs* in the CC condition based on single cues did not reveal significant effects ( $p > .05$ ).

We then tested whether fMRI-based *cs* in the IC condition could be predicted from single cue data. Here, regression analysis used single cue fMRI-based *cs* as predictors with the fMRI-based *cs* associated with the IC cue as a dependent variable. The relationship proved significant in the rDIPS ( $r^2 = .593$ ,  $\beta_{\text{shading}} = .447$ ,  $\beta_{\text{motion}} = .552$ ,  $F_{(2,10)} = 5.835$ ,  $p = .027$ ) (Fig. 4F). While similar trends towards a linear combination of single curvature sensitivities in the combined cue conditions were also observed in IMT ( $p = .06$ ), (for CC curvature) as well as in IDIPS ( $p = .081$ ) (for IC curvature), regression analysis in the other ROIs did not reach significance.

Hence, our data in the CC condition show that fMRI-based *cs* in rVIPS and rPPA can be predicted by a linear combination of *cs* obtained in single cue conditions. In Figure 4C this means that in case shading and motion conditions show different signs of *cs* (e.g. see participant AL), the *cs* in the CC condition has a tendency to be a weighted average of single cues *cs*. In case both single cues have the same sign of *cs*, the CC condition tends to adopt the sign of the single cue *cs* (either increase or decrease in beta values as a function of curvature) (e.g. see participant VK). These relationships between single cue *cs* and CC cue *cs* are re-plotted in Figure 4E, showing the associated regression surfaces. Single dots represent the *cs* (slopes) of single participants in each of the 3 conditions in 3D space.

Likewise, our data in the IC condition showed that in the rDIPS, a linear combination of *cs* in single cue conditions predicted *cs* in the IC condition. In Figure 4D, this means that the *cs* in the IC condition has a tendency to be a weighted average of single cue *cs* irrespective of single cues showing opposite or same signs of *cs* (e.g. participant AL and participant FD). The associated regression surfaces are shown in Figure 4F.

Interestingly, systematic relationships between participants' single cue fMRI-based *cs* and combined cue fMRI-based *cs* (slopes) in CC and IC conditions were only found in right hemispheric regions (i.e. rPPA, rVIPS, rDIPS) but not in their left hemispheric equivalents.



**Fig. 4** Single cue fMRI-based *cs* predicts combined cue fMRI-based *cs* in high-level right hemispheric regions. **A&B** Group averaged fMRI-based *cs* in rVIPS and rDIPS. Beta values in the group average data showed no significant modulation by stimulus curvature in any of the 4 cue conditions (yellow: shading, blue: motion, purple: CC, green: IC). **C&D** Single participant fMRI-based *cs* in rVIPS and rDIPS. *cs* was quantified by the slope of linear trend lines fit to beta values as a function of curvature. Data points in the main plots represent the single cue and CC *cs* in rVIPS (panel C) as well as the single cue and IC *cs* in rDIPS (panel D) for each of the 11 participants. For visualization purposes data points of individual participants in the same cue condition are connected by a dotted line. Data points are ordered from left to right according to the slope values (i.e. curvature sensitivities) in the single shading cue. *cs* could increase (positive slope) or decrease (negative slope) as a function of curvature and were highly variable across cues and participants (see insets for examples of single subject beta values). **E&F** Single cue fMRI-based *cs* predicts combined cue fMRI-based *cs*. Regression surfaces associated with a multiple linear regression analysis using *cs* in single cues to predict *cs* in combined cues. In rVIPS, single shading and motion fMRI-based *cs* significantly predicted CC cue fMRI-based *cs* (panel E). Similarly, in rDIPS, single shading and motion fMRI-based *cs* significantly predicted IC cue fMRI-based *cs* (panel F). Dots represent the slopes of each participant in the 3 cue conditions in 3D space. For visualization purposes a dot's color indicates the position in the upward direction in correspondence with the color scheme of the regression surface (with red-to-blue values corresponding to respectively positive-to-negative *cs* values). Additionally, black lines indicate a dot's position on the shading and motion axes. rVIPS = right ventral intraparietal sulcus, rDIPS = right dorsal intraparietal sulcus.

### **Relationship of performance data with fMRI-based curvature sensitivity**

Since regions rPPA, rVIPS and rDIPS revealed curvature-dependent activation on an individual participant level, we here tested for a correlation between fMRI-based *cs* and performance-based *cs* in these ROIs. Recall that performance-based *cs* corresponded to the slopes of linear trend lines fitted to proportions of ‘larger’ responses as a function of curvature. Similarly, fMRI-based *cs* corresponded to the slopes of linear trend lines fitted to beta values plotted as a function of curvature. Remarkably, results revealed that performance-based *cs* and fMRI-based *cs* of the same cue type did not correlate significantly in any of the ROIs ( $p > .05$ ). This was also true for correlations of performance-based *cs* with fMRI-based *cs* using their absolute values (irrespective of whether beta values increased or decreased as a function of curvature) ( $p > .05$ ).

Since in the above reported analysis a relationship between single cue fMRI-based *cs* and combined cue fMRI-based *cs* was found in rPPA, rVIPS and rDIPS, we here additionally investigated whether combinations of single cue fMRI-based *cs* in the same regions were also predictive of psychophysical performance in CC and IC cue conditions. However, our regression analysis using fMRI-based *cs* of single cues as independent variables did not predict performance-based *cs* in neither of the combined cue conditions ( $p > .05$ ).

In addition to the above reported regression analysis, we also computed fMRI-based *cs* cue weights to further investigate the relationship between combinations of single cue fMRI-based *cs* and psychophysical performance in combined cue conditions. First, the value of the fMRI-based *cs* cue weights was determined by the absolute difference between shading and motion fMRI-based *cs*. Then the sign of this difference value (positive or negative) was determined by the cue with the larger absolute slope value (i.e. the larger *cs*). Stronger sensitivity to shading-defined curvature (larger shading cue *cs*) was indicated by positive difference values and stronger sensitivity to motion-defined curvature (larger motion cue *cs*) was indicated by negative difference values (see Methods). These fMRI-based *cs* cue weights were then correlated with performance-based *cs* in the CC and IC conditions. While there was no significant correlation in any of the 3 ROIs between fMRI-based *cs* cue weights and performance-based *cs* in the CC condition, interestingly, fMRI-based *cs* cue weights derived from rDIPS showed a (weak) correlation with performance-based *cs* in the IC condition ( $r = .526$ ,  $p = .048$ ). Figure 5A shows that positive fMRI-based *cs* cue weights (sensitivity to shading-derived curvature > sensitivity to motion-derived curvature) were associated with shading-dominated performance in the IC condition (steep positive performance-based *cs*). Moreover, negative fMRI-based *cs* cue weights (sensitivity to shading-derived curvature < sensitivity to motion-derived curvature) were associated with motion-dominated performance (negative performance-based *cs*). Small or close-to-zero *cs* weights indicate approximately equal sensitivity to both single cues whereas large values indicate a large difference in fMRI-based *cs* between single cues. Note that our cue weight estimation based on performance data revealed a slightly larger weighting of the shading cue in the IC condition (Fig. 2B), which explains the scarcity of negative (motion-dominant) behavioural slopes. A similar

relationship between fMRI-based *cs* cue weights and performance in the IC condition was also observed in rPPA which however missed significance ( $p = .065$ ). Thus, in rDIPS fMRI-based *cs* cue weights derived from single cues predicted performance-based *cs* in the IC condition. IC condition performance in single participants was dominated by the single cue that showed a larger fMRI-based *cs*.

In sum, in individual participants, there were no significant relationships between fMRI-based *cs* and performance-based *cs* in the same cue condition, making these observations difficult to interpret. However, the finding of a predictable relationship between single and combined cue fMRI-based *cs* in rDIPS was supported by a correlation (albeit weak) between single cue fMRI-based *cs* cue weights and performance-based *cs* in the IC condition.

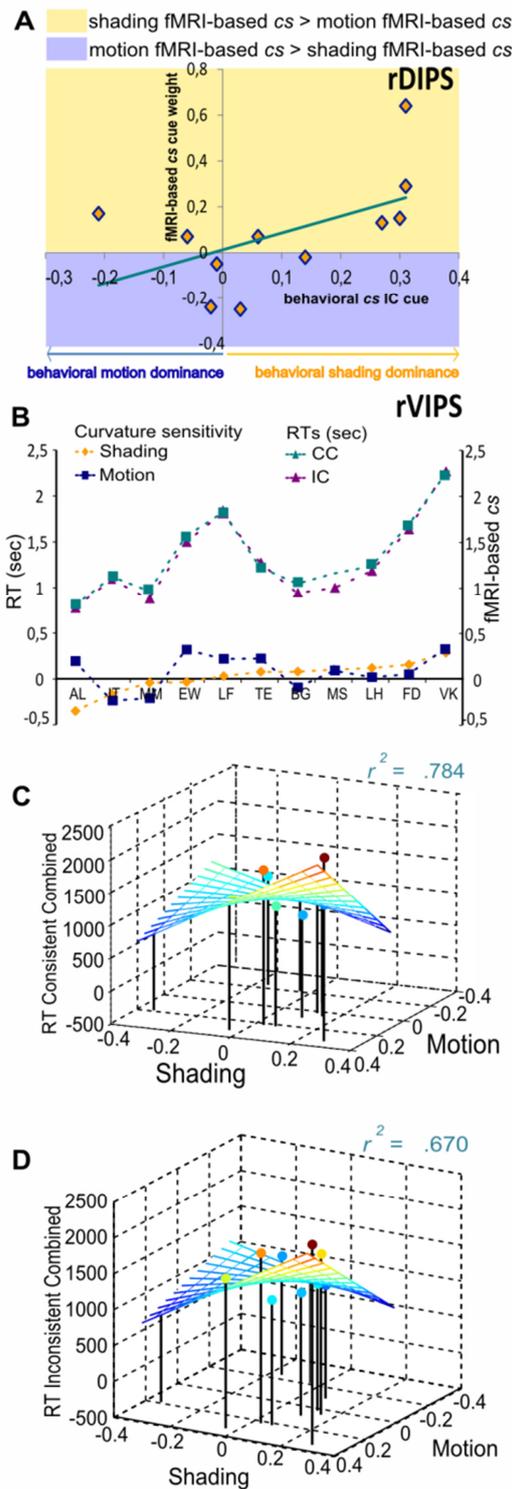
### ***Relationship of RTs with fMRI-based curvature sensitivity***

Similar to the analysis in the previous paragraph in which we intended to establish a correlation with performance data and fMRI-based *cs*, we also tested for a correlation between RTs and fMRI-based *cs* in rPPA, rVIPS and rDIPS.

Again, within each single cue condition, RTs and fMRI-based *cs* did not correlate in any of the ROIs ( $p > .05$ ). However, based on the previous finding of a relationship between single cue fMRI-based *cs* and combined cue fMRI-based *cs*, we here intended to test whether in a multiple linear regression analysis fMRI-based *cs* from single cues in the same right hemispheric regions (rVIPS, rPPA and rDIPS) could also predict RTs in combined cue conditions. In Figure 5B, the relationship between fMRI-based *cs* of single cues in rVIPS and RTs are plotted for each of the 11 participants. Concerning the CC condition, RTs tended to be longer in participants who displayed fMRI-based *cs* that had *cs*-values (slopes) of the same sign for both shading and motion cues, whereas RTs tended to be shorter in participants with opposing slope signs for shading and motion cues (increase as a function of curvature for one cue and decrease for the other). For example, participant AL showed a positive slope for motion-defined curvature, a negative slope for shading-defined curvature and short RTs in combined cue conditions. Participant VK showed an increase in beta values as a function of curvature defined by both cues (positive slope) and had longer RTs in combined cue conditions (Fig. 5B). A relationship between single cue fMRI-based *cs* and CC condition RTs was confirmed in a logistic regression analysis in area rVIPS (Fig. 5C) ( $r^2 = .784$ ,  $\beta_{shading} = 334,876$ ,  $\beta_{motion} = 1287.212$ ,  $\beta_{shading \times motion} = 6685.329$ ,  $F_{(3,7)} = 8.486$ ,  $p = .01$ ).

In the IC condition, as the RTs in IC and CC conditions were similar within participants (Fig. 5B), multiple regression analysis analogous to that performed for the CC, also revealed significant in the IC condition, yielding similar 3D regression planes ( $r^2 = .670$ ,  $\beta_{shading} = 67.875$ ,  $\beta_{motion} = 1302.290$ ,  $\beta_{shading \times motion} = 7838.583$ ,  $F_{(3,7)} = 4.730$ ,  $p = .042$ ). Figure 5D shows the results of the regression visually. Regression analyses in rPPA and rDIPS did not show any statistically reliable effects ( $p > .05$ ).

This analysis appears to suggest that the presence of opposite relationships between single-cue fMRI activity and curvature in right IPS causes RTs to decrease when a combined cue stimulus is presented, irrespective of whether the stimulus contains a consistent or inconsistent definition of curvature. The fact however that fMRI-based  $cs$  in a respective stimulus condition was uncorrelated with RTs in the same condition makes this finding hard to interpret.

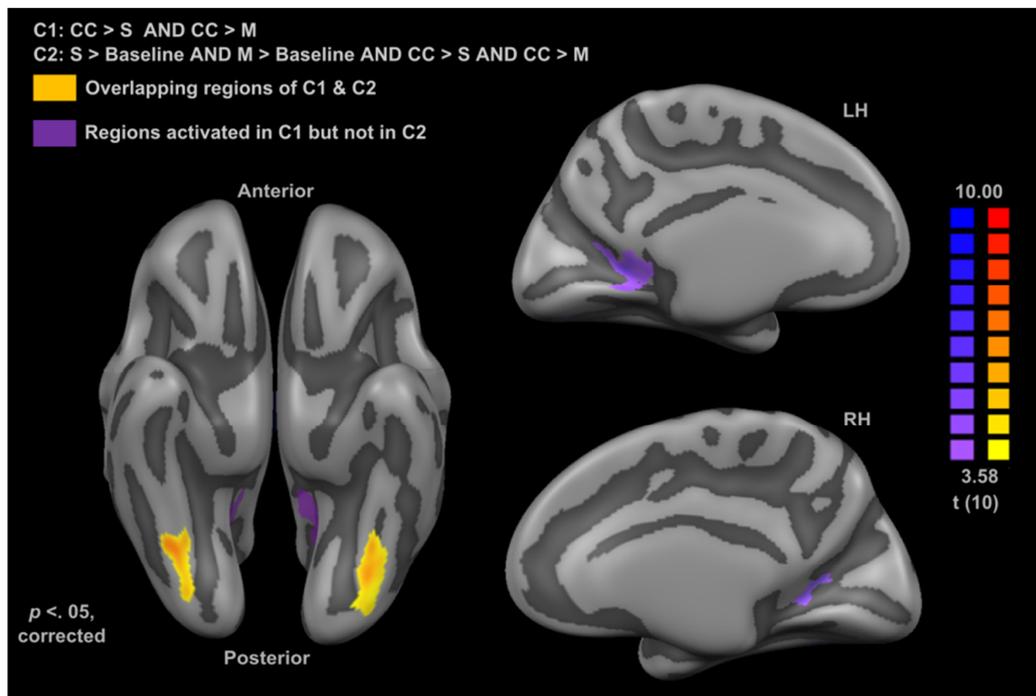


**Fig. 5 Single cue fMRI-based  $cs$  correlates with psychophysically derived measures** **A** Performance-based  $cs$  and fMRI-based  $cs$  cue weights. fMRI-based  $cs$  cue weights in rDIPS, reflecting the preference for coding shading- or motion-defined curvatures, significantly correlated with task performance in the IC condition. Positive fMRI-based  $cs$  weights signify larger sensitivity to shading- than to motion-defined curvature (indicated by yellow background coloring). Negative fMRI-based  $cs$  weights signify larger sensitivity to motion than to shading-defined curvature (indicated by blue background coloring). The value of the fMRI-based  $cs$  cue weights indicates the absolute difference between fMRI-based  $cs$  in single cues. Participants with positive fMRI-based  $cs$  cue weights revealed performance in the IC cue condition that was determined by the shading cue (positive values on the x-axis indicate behavioral shading-dominance). Participants with negative fMRI-based  $cs$  cue weights showed performance with stronger weighting of the motion cue in the IC curvature percept (negative values on the x-axis indicate behavioral motion-dominance) **B** Reaction times and fMRI-based  $cs$ . The relationship between fMRI-based  $cs$  in single cue conditions in rVIPS and RTs in CC and IC conditions is plotted for each of the 11 participants. RTs tended to be longer in participants who displayed fMRI-based  $cs$  to both shading and motion cues in the same direction (either an increase or a decrease as a function of curvature) whereas RTs tended to be shorter in participants with opposing slope signs for shading and motion cues (increase as a function of curvature for one cue and decrease for the other). For visualization purposes data points of individual participants in the same cue condition are connected by a dotted line. Data points are ordered from left to right according to the curvature sensitivities (i.e. slope values) in the single shading condition. **C&D** RTs in consistent and inconsistent combined conditions were significantly predicted by a combination of single cue fMRI-based  $cs$  in rVIPS. The resulting regression surfaces are shown. Single dots represent each of the 3 measures in 3D space for each participant. For visualization purposes a dot's color indicates the position in the upward direction in correspondence with the color scheme of the regression surface. Additionally, black lines indicate a dot's position on the shading and motion axes.

## Whole brain analysis

The conjunction contrast map of CC > shading and CC > motion (map C1 in Fig. 5) revealed bilateral regions in the posterior fusiform gyrus (color-coded in yellow) as well as bilateral activation at the junction of the occipito-temporal fissure and the calcarine sulcus (color-coded in purple). This shows that these regions responded generally more strongly to consistent combined stimuli than to each single cue alone. The conjunction contrast C1 was further constrained by excluding regions that were not significantly activated by each single cue alone (thus requiring for shading and motion activity to be above baseline; single cues > baseline). This second conjunction contrast (C2) yielded activation in the posterior fusiform gyrus (same as in conjunction contrast C1, color-coded in yellow), but not in the occipito-temporal fissure/calcarine sulcus region (Fig. 5). This indicates that the posterior fusiform gyrus showed *more* activation in combined cues than in single cues whereas the occipito-temporal fissure/calcarine sulcus region was activated only by combined stimuli and did not respond to each cue in isolation.

Additionally, we computed the contrast consistent > inconsistent cues. Areas in the posterior insula, the ACC (and several other frontal regions), the anterior STS as well as regions of the pre- and postcentral gyrus showed more activation to consistent combined conditions than to inconsistent combined conditions. Regions showing the opposite pattern were not found.



**Fig. 6 Regions coding combined shading-motion percepts.** Surface map resulting from conjunction contrasts overlaid on cortex-based aligned group average meshes ( $p < .05$ , corrected). Regions with larger activation in CC cue conditions than in single cues (yellow and purple color coding) were located in the bilateral posterior fusiform gyrus and at the junction of occipito-temporal fissure and calcarine sulcus. The purple color-coding indicates that the latter region was selectively activated by CC cues but not by single cues vs baseline (revealed in C1 but not in C2).

## Discussion

In the present study, we used fMRI in combination with a parametric psychophysical design to examine the mechanisms mediating shading and motion visual cue combination. Performance and RTs in combined cue conditions were a weighted average of single cue measures. However, variations in behavioural discrimination performance as a function of stimulus curvature (performance-based curvature sensitivity (*cs*), see Methods) did not lead to corresponding changes in fMRI activity (fMRI-based *cs*), neither in lower-level nor in higher-level visual areas. Despite the absence of a systematic signal variation as a function of curvature in group averaged data, individual participants did show strong relationships between fMRI activity and stimulus curvature. Surprisingly, fMRI activity to single or combined cues could increase or decrease as a function of curvature (i.e. a positive or negative signed fMRI-based *cs*) and was highly variable across cues and participants. However, our results show that, on an individual basis, fMRI-based *cs* in combined cue conditions in rIPS and rPPA were predicted by a weighted combination of fMRI-based *cs* in single cue conditions. In addition we found that some aspect of performance in combined cue conditions could be predicted from a combination of single cue fMRI-based *cs*, but not directly from fMRI activity yielded by the combined stimuli. Below we discuss the insights from preliminary behavioural and fMRI data analyses as well as avenues for further analysis and research.

### **Psychophysical data: Weighted single cue averaging in combined cue conditions**

Computational work suggests that combined cues produce more reliable estimates than single cues (Ernst and Banks 2002) and psychophysical studies for several combinations of (visual) depth cues have shown that human observers combine cues in a statistically optimal fashion, resulting in greater perceptual sensitivity (Young et al. 1993; Hillis et al. 2004). Previously we have shown (chapter 1) that the specific combination of shading and motion cues is not governed by MLE but instead yielded combined cue estimates that were *less* reliable than estimates based on their single cue constituents. However, overall performance in combined cue conditions was consistent with a weighted averaging of performance in individual cues. In the present study, we were limited to 4 curvature levels, which precluded the fitting of logistic response functions and a detailed assessment of reliability/sensitivity in the behavioral data. Nevertheless, in line with our previous findings (chapter 1), data in the present study also showed support for a weighted averaging of single cue response functions in combined cue conditions. Interestingly, the weight of the motion cue was relatively small in the consistent condition, but increased in the inconsistent condition. Note that in our previous study, despite a small increase in the motion cue weight in the inconsistent condition, shading remained very dominant. Here, due to the much larger cue conflict, the increase in the motion cue weight in the inconsistent combined

conditions compared to the consistent combined conditions was much more pronounced. This suggests a relationship between cue weighting and stimulus inconsistencies in which larger stimulus conflict results in a stronger weighting of the weaker cue.

### **fMRI data: Exploring the relationship between stimulus curvature and fMRI signal**

In the group average no consistent dependencies of fMRI activation in response to stimulus curvature were found. Nevertheless, at the level of individual participants, we did find fMRI activation in two right-lateralized IPS regions as well as in the right PPA that showed such curvature dependencies. However, the relationship between curvature and fMRI signal was highly variable across cues and participants. Since IPS has additionally been shown to contain a generic, cue-invariant representation of depth (depth-from shading: Taira et al. 2001; depth-from-motion: Kriegeskorte et al. 2003) and given our interest in cue combinations, we set out to investigate two questions: First, is the fMRI signal dependent on curvature in single cue conditions and how does psychophysical performance relate to this? Second, is the fMRI signal in single cue conditions predictive of the fMRI signal in combined cue conditions, and does the signal variation relate to psychophysical performance?

Regarding the first question, our analysis started from the idea that neuronal population responses in our localized ROIs would vary as a function of stimulus curvature. If curvature behaves as a form of 3D contrast in which a figure is more or less well defined against a background, we reasoned that this might be true at least in higher-level areas related to object processing (Graewe et al. 2012). However, since individual data did not show a relationship between performance and fMRI-based *cs* in single cue conditions (independent of whether *cs* was witnessed by an increase or decrease in fMRI signal as a function of curvature), it appears that curvature in the simple objects we used was not coded in the brain as we had hypothesized. This absence of a relationship between single cue fMRI-based *cs* and behavioural performance in principle compromised the possibility that relationships would be found between performance and fMRI activity in combined cue conditions.

Concerning the second question, we found interesting relationships in the fMRI data between single and combined cue conditions within rVIPS, rDIPS and rPPA. In particular, the strength and sign (increase or decrease as a function of curvature) of the fMRI-based *cs* in single cues was directly related to the strength and sign of fMRI *cs* in the combined conditions. Note that in combined cue conditions, the sensitivity of the fMRI activation to curvature was not generally larger than in single cue conditions. This can be seen as in line with a weighted averaging of single cue fMRI-activation in combined cue conditions. Moreover, rVIPS and rDIPS showed differential relationships between single cue and combined cue fMRI-based *cs* in consistent and inconsistent stimulus conditions. Whereas in rVIPS (ventral part of the right IPS), single cue conditions predicted curvature-sensitive activation in the *consistent* combined condition, in rDIPS (dorsal part of

the right IPS) single cue conditions predicted curvature-dependent activation in the *inconsistent* combined condition. Interestingly, these predictions showed two correlations with performance measures. First, in rVIPS, we found a relationship between single cue fMRI-based *cs* and RT in both CC and IC conditions. Second, in rDiPS we found a relationship (although relatively weak) between the strength of fMRI-based curvature encoding in each single cue and performance-based *cs* in the IC condition. Remarkably, in both cases, a certain aspect of performance during presentation of combined cue stimuli can be predicted from a combination of single cues, and not directly from fMRI activity yielded by the combined stimuli. This is consistent with a lack of behavioral integration between the cues.

A role of IPS in the detection of perceptual inconsistencies has previously been shown by Meienbrock et al. (2007). Furthermore, both ROIs in which curvature-related activation appeared to have an impact on some aspect of psychophysical performance were right lateralized, consistent with a role of right IPS in (visuo-)spatial functioning (Mesulam 1999; Sack et al. 2002).

### **Tentative links with opposite tuning properties of single neurons**

Intriguingly, while in some participants increases in depth-from-shading and depth-from-motion were mirrored by parametrically *increased* fMRI responses, in other participants fMRI activity *increased* if curvature was defined by one cue but *decreased* if curvature was defined by the other cue. At first glance, this seems related to electrophysiological recordings in macaques in which some multisensory cells showed opposite tuning patterns for the same object property specified by different sensory cues (Gu et al. 2008). For example, a recorded cell might be sensitive to leftward motion from a visual cue but to rightward motion from a vestibular cue. Cells showing these incongruent preferences would show diminished activity under consistent combined cue conditions compared to cells with congruent tuning preferences. Additionally, responses of congruent cells would be more strongly correlated with perceptual decisions than responses of cells showing incongruent tuning. Thus, congruent cells might provide a basis for optimal integration of vestibular and visual cues shown in monkeys (Fetsch et al. 2012) and humans (Butler et al. 2010). In the present fMRI study, paralleling the behavior of single cells in Gu et al. (2008), individually localized ROIs showed curvature-dependent fMRI activation, which in some cases could be opposite for the two cues (increase vs decrease as a function of curvature). This finding may suggest that some curvature-related information is present in rIPS and PPA regions but that it may be represented in an *absolute* manner (i.e. irrespective of whether it is coded by an increase or a decrease of fMRI activation). However, here, we found no correlation between curvature-dependent fMRI activation and psychophysical performance. Since participants were performing a task on the stimuli, factors other than low-level stimulus parameters, related to attention and task performance strategies, might have been important determinants of the observed fMRI signal variations in different curvature conditions. However, this would not necessarily preclude a correlation of fMRI signal with performance.

## Limitations and future analysis approaches

Our GLM-based analysis has shown that the changes in fMRI activity as a function of curvature in single cue and combined cue conditions are not easy to relate in any straightforward manner to either stimulus parameters or behaviour. It is possible that 3D curvature in simple objects as used in the present study is represented in a multivariate way, similar to orientation. Hence, variations of the fMRI signal in response to low-level curvature changes may better be indicated by MVPA which has proven successful in decoding parameters such as orientation (Haynes and Rees 2005) but also disparity-defined depth (Preston et al. 2008). Therefore, by using MVPA in the present dataset, one could not only test the decoding of curvature-related information in single cue conditions, but also whether patterns of activity elicited by our combined cue stimuli would be predicted from weighted averaging of patterns of activity obtained in response to single cue stimuli (Reddy et al. 2009). Since different visual depth cues have been shown to be processed in overlapping regions in the IPS (Georgieva et al. 2008; Gardin et al. 2010; Orban 2011), analysis in this region might be interesting. In fact, two ROIs in the IPS were implicated in the preliminary analyses presented here as showing consistent relationships between activity induced by single and combined cue stimuli, although in a manner difficult to relate to behavioural performance.

Additionally, a potentially promising ROI to focus on in a follow-up study is a region at the junction of the occipito-temporal fissure and calcarine sulcus which was identified by our whole brain analysis as exclusively activated by combined but not by single cue stimuli. This region is in the vicinity of a retinotopically-defined region termed V6 (Pitzalis et al. 2006) and has previously been shown to be selective to other depth cue combinations such as optic flow patterns paired with consistent rather than inconsistent or zero-disparity gradients (Cardin and Smith 2011). Additionally, Vaina (1989) reported selective loss of structure-from-motion and stereopsis after lesions to (the vicinity of) this region.

As indicated in chapter 2, the lack of cooperative integration in the behavioural data may reflect 3D curvature ambiguity in our stimuli (Bulthoff and Mallot 1987; Orban 2011; Lovell and Bloj 2012). It would therefore be interesting to see whether in an experiment in which one or both of the cues are disambiguated (e.g. by using a fixed position of the light source in the shading stimulus or increasing the dot density of the motion stimulus), cooperative cue integration at the behavioural level would be found. In this case, an MVPA analysis might show that curvature-dependent activity patterns related to combined cues are more robust than those related to single cues, thereby increasing classification accuracy in a manner that would parallel cooperative integration during psychophysical curvature judgements.

## Conclusions

The current study advances our understanding of shading and motion cue combinations, but so far also leaves several questions unanswered. The weighted averaging of single cue performance and RTs in the psychophysical data show, in

line with previous findings (chapter 2), that MLE is an insufficient theoretical frame to explain our data. Instead a Bayesian approach, taking several ecological factors into account may be more appropriate. Moreover, weighted averaging of curvature-related single cue fMRI activation predicted curvature-related fMRI activation in high-level visual areas (rPPA, rVIPS and rDIPS) in combined cue conditions. However, there were no clear relationships between behavioral performance and the way in which single or combined cue curvatures were coded, making these findings difficult to interpret at this point.

Thus, our fMRI analysis so far has not revealed evidence related to the manner in which the cues are represented in the absence of cooperative integration, and the design of the experiment has neither allowed us to compare attended and non-attended conditions, nor to assess the effects of disambiguation of the stimuli on the modus of cue integration. Resolving these interesting questions will require further research and analysis.

## References

- Bakin JS, Nakayama K, Gilbert CD. 2000. Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J Neurosci.* 20:8188–8198.
- Ban H, Preston TJ, Meeson A, Welchman AE. 2012. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci.* 15:636–643.
- Brouwer GJ, Tong F, Hagoort P, van Ee R. 2009. Perceptual incongruence influences bistability and cortical activation. *PLoS One.* 4:e5056.
- Bulthoff H, Mallot H. 1987. Interaction of different modules in depth perception. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Technical report.
- Butler J, Smith S, Campos J, Bulthoff H. 2010. Bayesian integration of visual and vestibular signals for heading. *J Vis.* 10:1–13.
- Cardin V, Smith AT. 2011. Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion. *J Neurophysiol.* 106:1240–1249.
- Clark J, Yuille A. 1990. Data fusion for sensory information processing systems. Kluwer Academic Publishers.
- Dien J. 2009. A tale of two recognition systems: implications of the fusiform face area and the visual word form area for lateralized object recognition models. *Neuropsychologia.* 47:1–16.
- Dupont P, De Bruyn B, Vandenberghe R, Rosier a M, Michiels J, Marchal G, Mortelmans L, Orban G a. 1997. The kinetic occipital region in human visual cortex. *Cereb Cortex.* 7:283–292.
- Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature.* 415:429–433.
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. 2012. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci.* 15:146–154.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnet Reson Med.* 33:636–646.
- Georgieva SS, Todd JT, Peeters R, Orban G a. 2008. The extraction of 3D shape from texture and shading in the human brain. *Cereb Cortex.* 18:2416–2438.
- Gerardin P, Kourtzi Z, Mamassian P. 2010. Prior knowledge of illumination for 3D perception in the human brain. *P Natl Acad Sci USA.* 107:16309–16314.
- Goebel R, Esposito F, Formisano E. 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp.* 27:392–401.

- Goebel R, Khorrám-Sefat D, Muckli L, Hacker H, Singer W. 1998. The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur J Neurosci.* 10:1563–1573.
- Graewe B, Lemos R, Ferreira C, Santana I, Farivar R, De Weerd P, Castelo-Branco M. 2012. Impaired Processing of 3D Motion-Defined Faces in Mild Cognitive Impairment and Healthy Aging: An fMRI Study. *Cereb Cortex.* doi: 10.10.
- Gu Y, Angelaki DE, Deangelis GC. 2008. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci.* 11:1201–1210.
- Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci.* 8:686–691.
- Hillis J, Watt S, Landy M, Banks M. 2004. Slant from texture and disparity cues: Optimal cue combination. *J Vis.* 4:967–992.
- Hillis JM, Ernst MO, Banks MS, Landy MS. 2002. Combining sensory information: mandatory fusion within, but not between, senses. *Science.* 298:1627–1630.
- Jacobs R, Fine I. 1999. Experience-dependent integration of texture and motion cues to depth. *Vision Res.* 39:4062–4075.
- Johnston EB, Cumming BG, Landy MS. 1994. Integration of stereopsis and motion shape cues. *Vision Res.* 34:2259–2275.
- Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. *J Neurosci.* 23:1451–1463.
- Liu Y, Vogels R, Orban G a. 2004. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *J Neurosci.* 24:3795–3800.
- Lovell P, Bloj M. 2012. Optimal integration of shading and binocular disparity for depth perception. *J Vis.* 12:1–18.
- Meienbrock a, Naumer MJ, Doehrmann O, Singer W, Muckli L. 2007. Retinotopic effects during spatial audio-visual integration. *Neuropsychologia.* 45:531–539.
- Mesulam MM. 1999. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos T Roy Soc B.* 354:1325–1346.
- Orban G a. 2011. The extraction of 3D shape in the visual system of human and nonhuman primates. *Annu Rev Neurosci.* 34:361–388.
- Pitzalis S, Galletti C, Huang R-S, Patria F, Committeri G, Galati G, Fattori P, Sereno MI. 2006. Wide-field retinotopy defines human cortical visual area v6. *J Neurosci.* 26:7962–7973.
- Preston TJ, Li S, Kourtzi Z, Welchman AE. 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J Neurosci.* 28:11315–11327.

- Qiu F, Heydt RVD. 2005. Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron*. 47:155–166.
- Reddy L, Kanwisher N, Van Rullen R. 2009. Attention and biased competition in multi-voxel object representations. *P Natl Acad Sci USA*. 106:21447–21452.
- Sack A, Sperling J, Prvulovic D. 2002. Tracking the Mind's Image in the Brain II-Transcranial Magnetic Stimulation Reveals Parietal Asymmetry in Visuospatial Imagery. *Neuron*. 35:195–204.
- Schiller PH, Slocum WM, Jao B, Weiner VS. 2011. The integration of disparity, shading and motion parallax cues for depth perception in humans and monkeys. *Brain Res*. 1377:67–77.
- Sereno MI, Dale a M, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*. 268:889–893.
- Taira M, Nose I, Inoue K, Tsutsui K. 2001. Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. *NeuroImage*. 14:959–966.
- Vaina L. 1989. Selective impairment of visual motion interpretation following lesions of the right occipito-parietal area in humans. *Biol Cybern*. 359:347–359.
- Welchman AE, Deubelius A, Conrad V, Bülhoff HH, Kourtzi Z. 2005. 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci*. 8:820–827.
- Young MJ, Landy MS, Maloney LT. 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res*. 33:2685–2696.
- van Atteveldt N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. *Neuron*. 43:271–282.

## Stimulus Dependency of Object-Evoked Responses in Human Visual Cortex: An Inverse Problem for Category Specificity

---

*Corresponding publication:* Graewe B, De Weerd P, Farivar R, Castelo-Branco M (2012) Stimulus Dependency of Object-Evoked Responses in Human Visual Cortex: An Inverse Problem for Category Specificity. PLoS ONE 7(2): e30727. doi:10.1371/journal.pone.0030727

## Abstract

Many studies have linked the processing of different object categories to specific event-related potentials (ERPs) such as the face-specific N170. Despite reports showing that object-related ERPs are influenced by visual stimulus features, there is consensus that these components primarily reflect categorical aspects of the stimuli. Here, we re-investigated this idea by systematically measuring the effects of visual feature manipulations on ERP responses elicited by both structure-from-motion (SFM)-defined and luminance-defined object stimuli. SFM objects elicited a novel component at 200-250 ms (N250) over parietal and posterior temporal sites. We found, however, that the N250 amplitude was unaffected by restructuring SFM stimuli into meaningless objects based on identical visual cues. This suggests that this N250 peak was not uniquely linked to categorical aspects of the objects, but is strongly determined by visual stimulus features. We provide strong support for this hypothesis by parametrically manipulating the depth range of both SFM- and luminance-defined object stimuli and showing that the N250 evoked by SFM stimuli as well as the well-known N170 to static faces were sensitive to this manipulation. Importantly, this effect could not be attributed to compromised object categorization in low depth stimuli, confirming a strong impact of visual stimulus features on object-related ERP signals. As ERP components linked with visual categorical object perception are likely determined by multiple stimulus features, this creates an interesting inverse problem when deriving specific perceptual processes from variations in the amplitude of ERP components.

## Introduction

There is a strong ongoing debate on the category specificity of event-related potentials (ERPs). Especially in the case of face stimuli, many studies have interpreted their results in favour of category-specific processing. While differences in amplitude related to object categories have been reported as early as 100 ms (P100) (Herrmann et al. 2005) and in later components such as the N250 (Schweinberger et al. 2004), the 'N170', a negative potential over occipital-temporal sites is the most famous and well-studied peak showing a larger amplitude for faces than for other object categories (Bentin et al. 1996; Eimer 2000; Itier 2004; Itier and Taylor 2004). Similarly, the discovery of the face-inversion effect demonstrating a delayed and enhanced N170 to inverted faces but not to inverted objects (Rossion et al. 1999, 2000) has contributed to a consensus on the face specificity of this peak. Stimuli in prior studies of category-specific ERP components typically consisted of photographs (in black and white, grey-levels or color) or schematized images (line drawings, cartoons or Mooney stimuli) (for a review see Rossion and Jacques 2011). However, even if the dimensions that are represented in high-level cortex cannot be reduced to physical features (Kayaert et al. 2005; Haushofer et al. 2008), low-level physical properties unavoidably differ between categories and may therefore contribute to differential ERP responses (Troje and Bühlhoff 1996; VanRullen and Thorpe 2001; Rousselet, Husk, et al. 2008; Rousselet, Pernet, et al. 2008). Indeed, in the case of the N170, many results characterize a sensitivity to a variety of stimulus manipulations including *feature isolation* (Bentin et al. 1996), *contrast polarity reversal* (Itier and Taylor 2002), *spatial frequency filtering* (Goffaux et al. 2003; Vlamings et al. 2009) and *stimulus orientation* (Jacques and Rossion 2007). While some studies have equated images on global factors such as contrast, luminance level and spatial frequency (Rousselet, Husk, et al. 2008; Vizioli et al. 2010), different object categories have often been compared without using control images that were equated on visual features (Eimer and McCarthy 1999; Eimer 2000; Itier 2004) or by using (scrambled) control images that not only abolished categorical object information but also cue configurations and visual structure (Bentin et al. 1996; Allison et al. 1999; Latinus and Taylor 2006). In all of these studies, visual feature properties may have added significantly to the results.

To test the relative contributions of categorical and visual feature variations on ERP responses, we performed two types of experiments. First, we aimed to separate the impact of visual feature- and category-related factors on the ERP signal by comparing ERP peaks elicited by structure-from-motion-(SFM) objects and restructured control stimuli containing the same visual feature information (luminance, contrast, SF, motion) (Exp. 1A&B). In SFM stimuli the object percept arises solely from a moving dot pattern with the object being invisible when the dots are stationary. The use of dynamic localized cues (Farivar et al. 2009) avoids contour and other artifacts that would have appeared in restructured, static

luminance-defined control images hence making SFM stimuli well-suited for the purpose of the present study. Second, to investigate the impact of visual feature variations on object-related ERPs parametrically, we manipulated the stimulus depth range in SFM-defined and luminance-defined faces and chairs (Exp. 2B&C). The use of objects defined by both motion and luminance cues enabled us to strengthen the generality of our findings.

## Methods

### Participants

All participants ( $N = 48$ , ages 22-31, 22 males and 26 females) had normal or corrected-to normal vision and no history of neurological disorders. The experiment was approved by the local Ethics Committee of the University of Coimbra and before the start of the experiment written informed consent was obtained from all participants in accordance with the Helsinki Declaration. A total of 27 participants took part in EEG recordings, distributed over 4 experiments (Exp. 1A, 1B, 2B, 2C). The remaining 21 participants performed a psychophysical control experiment (Exp. 2A).

### Procedure

Participants were seated in a dimly lit sound-attenuated cabin viewing stimuli subtending  $\sim 13^\circ$  of visual angle horizontally and  $\sim 10^\circ$  vertically. Viewing distance was 80 cm and stimuli were always presented at the center of the screen of a CRT monitor (resolution 1024\*768, refresh rate 60Hz). Stimuli were delivered using Presentation 12.1 (Neurobehavioral systems).

### Experiment 1: Categorical stimuli versus matched control stimuli

The ERP responses evoked by SFM categorical and control stimuli were investigated in 14 participants, in two separate but comparable experiments (Experiment 1A: 8 participants, 5 female, 3 male; Experiment 1B: 6 participants, 3 female, 3 male).

#### **Experiment 1A**

##### *Stimuli*

SFM stimuli were videos of SFM-defined faces, chairs, and control stimuli (for details see Farivar et al. 2009). Face stimuli consisted of 10 laser-scanned facial surfaces taken from the Max-Planck Face Database, rendered with a volumetric texture map ensuring uniform texture density across the surface (as described in detail in Liu et al. 2005). Shadows and shading were removed and the faces were rendered against a similarly textured random-dot background. During stimulus presentation, the face rotated from  $-22.5$  degrees to  $22.5$  degrees centered at the frontal plane in one cycle (Fig. 1A) with the rotation being captured in a video that lasted 860 ms (26 frames). The 10 chair stimuli were obtained from a chair model

database and were rendered in exactly the same manner as the faces. Control versions of the two stimuli were constructed by cutting the rendered whole object (face or chair) videos in the horizontal plane into ten blocks and restructuring their positions within the object boundaries. Importantly, this manipulation did not introduce localized cues (high spatial frequency (HSF) noise confounds) but resulted in control stimuli that share many of the visual features of the original videos (including luminance, contrast, texture, spatial frequency, motion, and strength of depth and curvature cues in contours and surfaces). Thus, intact and control images were equated in visual features, with the intact stimuli showing a recognizable object, and the restructured control images a meaningless object entity.

### *Task*

It has previously been shown that for tasks in which participants attend to face identity, the N170 amplitude is enhanced for familiar compared to unfamiliar faces (Caharel et al. 2002, 2005). Participants were therefore trained to associate names with 10 SFM faces and 10 SFM chairs prior to the start of the EEG experiment. Training was halted when at least 80% of the responses were correct. To ensure sustained attention and task engagement during the EEG recordings, participants were asked to maintain eye fixation and were presented with a stimulus sequence consisting of an SFM stimulus (860 ms), a fixation period (1 sec) and the presentation of a name from the respective category (face or chair name; 1 sec). In the 1 sec response period following each trial, they then indicated by a button press whether according to the face-name/chair-name encoding they had done in the previous training, the presented name corresponded to the SFM stimulus (left button: yes, right button: no). Participants were asked to respond as accurately as possible and did not receive feedback on their responses. For consistency, button presses were also given in the control SFM stimulus conditions for which faces or chairs were however unidentifiable, hence resulting in random responses. The experiment was divided into 5 self-initiated runs of 12 minutes each. There were 250 trials for each stimulus category resulting in a total of 1000 trials which were presented in random order.

## **Experiment 1B**

### *Stimuli*

Here, the SFM stimuli used in experiment 1A (face, chair, and associated control stimuli) were shortened to 160 ms (5 frames) and rotated from -4.3 degrees to 4.3 degrees. Prior to the EEG recordings we created stimuli of 2 different durations (100 ms and 160 ms) and tested the effect of duration on psychophysical task performance.

### *Task*

Participants were asked to perform a simple categorization task in which they indicated by 3 alternative button presses whether the SFM stimulus contained a

face, chair or a meaningless object. Thus, in this task object-related processes that might have infiltrated ERP responses to restructured control stimuli (e.g., imagery) are minimized, both because of the unspecific categorization task that no longer required identification of individual exemplars, and because of the explicit categorization of the restructured control stimuli as ‘meaningless’. The two changes compared to experiment 1A (shortened duration and categorization task) aimed to reveal a larger contribution of bottom-up processing and to minimize effects of high-level cognitive factors. The response period was increased to 1.5 s, relaxing the time pressure for responding.

## **Experiment 2: Parametric variation of visual features in categorical stimuli**

In experiment 2A, there were 21 participants (11 female, 10 male); in experiment 2B, 8 participants (5 female, 3 male); and in experiment 2C, 5 participants (2 female, 3 male).

### ***Experiment 2A***

#### *Stimuli*

Depth range was varied for a single SFM face and a single SFM chair stimulus (duration 160 ms, see exp. 1 for details on the stimuli). This global manipulation of the elementary variable ‘depth’ resulted in new stimuli (parameterized in terms of anterior-posterior range) that were randomly presented at 3 levels of depth (10%, 30%, 90%, expressed as a percentage of the un-manipulated depth range). The change in depth caused variations in complexity and strength of a number of visual features, as can be seen in analogous depth-related variations in luminance-defined stimuli in Figure 1B. For simplicity, this manipulation will from hereon be referred to as ‘depth range manipulation’.

#### *Task*

Participants performed a psychophysical task (without EEG measurements) in which categorization performance was tested under varying depth levels of the SFM stimuli. Participants were instructed to distinguish between SFM face and chair stimuli by pressing one of two buttons in a 1.5 sec response period. There were 10 trials per condition.

### ***Experiment 2B***

#### *Stimuli*

As in experiment 2A, depth range was varied for a single SFM face and a single SFM chair stimulus (duration 160 ms, see exp. 1 for details on the stimuli). Here, we used 4 different depth levels (10%, 30%, 60% and 90% of the un-manipulated depth range).

#### *Task during EEG recording*

Similar to the psychophysical control task (Exp. 2A), single SFM faces or single SFM chairs were displayed in random order at the 4 depth levels. There were 120 trials

per condition with participants indicating by button press in a 1.5 s response period which stimulus category they had perceived (face or chair).

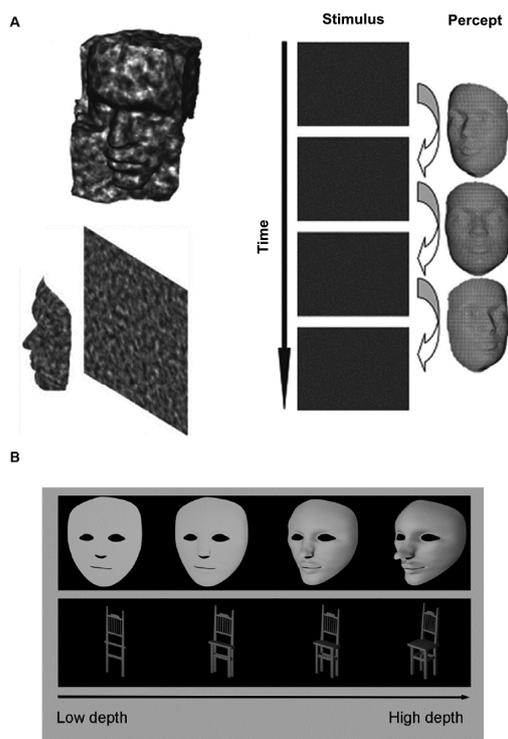
## Experiment 2C

### Stimuli

Luminance-defined static face and chair stimuli (Fig. 1B) were constructed using Blender software. The resulting monochromatic meshes were depth-manipulated along the y axis which in the face meshes implied variations in the convexity of facial features, while in the chair meshes it caused variations in the length of the seat and in the width of the legs and backrest. Thus, as was the case for SFM stimuli, this parametric manipulation varied the richness, strength and complexity of the stimulus/object cues. In the absence of changes in a participant's ability to categorize the stimuli, changes in ERP responses are likely to be driven by variations in one or more of such visual features, which would be informative for the question whether responses to object stimuli are driven by visual feature information or by categorical aspects. By including the luminance-defined stimuli, we aimed to answer this question for both SFM and luminance-defined object stimuli. As in experiment 2B, we used 4 different depth levels (10%, 30%, 60% and 90%). Stimuli were displayed for 500 ms (~15 frames).

### Task during EEG recording

Participants indicated by button presses in a 1 sec response period which category they had perceived (face or chair). There were 120 trials per condition.



**Figure 1. Stimuli used in the present study.** **A** Illustration of a structure-from-motion object stimulus. The percept of a rotating face emerges from a moving dot pattern in the absence of other visual cues (used in Exp. 1A, 1B, 2A, 2B) **B** Static luminance-defined faces and chairs with varying depth ranges (used in Exp. 2C). 3D structure was manipulated by computing views of ‘flattened’ objects causing variations in complexity and strength of a number of visual features (see Materials and Methods). (Figure adapted from (Farivar et al. 2009)).

## **ERP Recording and Analysis**

Continuous EEG data for experiment 1 were recorded with a NeuroScan SynAmp system at a sampling rate of 2000 Hz from 64 scalp electrodes embedded in a NeuroScan Quik-Cap. Two EKG channels monitored heart rate and four bipolar facial electrodes, positioned on the outer canthi of each eye and in the inferior and superior areas of each orbit, monitored horizontal and vertical EOG (HEOG and VEOG), respectively. The impedance of each electrode was adjusted to less than 10 k $\Omega$ . The signal was amplified using Synamps Amplifiers and recordings were made using CPz as an online reference. Following a system upgrade, experiments 1B and 2B, C were conducted with a 128-channel Neuroscan system under equal conditions and settings.

### ***Data Analysis***

After recording, data were analyzed offline using the Scan 4.3 Edit Software. The continuous data files were downsampled to 500 Hz and digitally low- and highpass filtered (30Hz, 12dB/oct and 1Hz, 12dB/oct, respectively). Artifact rejection was performed automatically, removing epochs with amplitudes exceeding  $\pm 75 \mu\text{V}$  and by subsequent visual inspection. Eye movements were corrected for by the blink-noise-reduction option. Artifact free continuous data were then segmented into epochs ranging from -200 ms to 500 ms after stimulus onset for all conditions. The epochs were baseline corrected based on the time interval -200 ms before stimulus onset and subsequently averaged per experimental condition.

### ***Peak Analysis***

For each participant, 3 peaks were automatically detected and included in the analysis. P100 was determined as the most positive peak between 70 and 140 ms and N2 as the most negative peak between 140 and 200 ms (140-180 ms in the parietal cluster). At later latencies, analysis revealed a large negative peak at  $\sim 250$  ms (defined as most negative peak between 210-280 ms; 80-280 ms in the parietal cluster). For the static stimuli in experiment 2C, the N170 peak was defined as the most negative peak between 140 and 210 ms. Based on previous studies and our focus on visual object recognition, statistical analysis was restricted to occipital, posterior temporal and parietal regions. Since the P100 is not the focus of our interest here and analysis in the occipital cluster did not reveal any relevant results, the figures only display ERP results from the parietal and posterior temporal cluster.

For the amplitude and latency analysis, clusters of electrodes were defined, yielding a regional weighted average of the electrodes showing the largest amplitudes and hence optimally revealing the peaks of interest. The occipital cluster included electrodes O1 and O2 (in the 128-channel system 20,21,42,43,44,45 on the left, corresponding sites on the right), the posterior temporal cluster TP7, P7, P05, P07 over the left hemisphere and corresponding sites on the right (in the 128-channel system 17,18,19,20,21,22,23 on the left and corresponding sites on the right) and the parietal cluster P3 and P4 (in the 128-

channel system electrodes 39,40,47,48 on the left and corresponding sites on the right). For the statistical analysis, a repeated measures GLM was applied on the amplitudes and latencies of all components of interest (based on our research question, in exp. 2 only peak amplitude values were included in the analysis). Where applicable, p-values were corrected for non-sphericity using the Greenhouse-Geisser or Huynh-Feldt correction (for simplicity, the uncorrected degrees of freedom are presented) and p-values for multiple comparisons were Bonferroni-corrected.

### **Source localization**

Brain generators were estimated over the time period from -100 to 500 ms using a distributed, linear solution to the inverse problem, based on the sLORETA method (standardized low resolution brain electromagnetic tomography (Pascual-Marqui et al. 1994)). This method has been shown to have no localization bias in the presence of measurement and biological noise (Pascual-Marqui 2002) and takes several neurophysiologic and anatomical constraints into account. Current source density maps were constructed assuming multiple simultaneously active sources of a priori unknown location and making no assumption regarding the number or location of active sources. For the source estimation, individual MR data were used to create a Boundary element (BEM) model. Functional ERP data and anatomical MRI scans of individual participants were co-registered using landmarks and applying standard xyz coordinates of the electrode positions on the head. After pre-processing the EEG data, an independent component analysis (ICA) was applied and only the main components (signal to noise ratio SNR > 1) were used for the source reconstruction. LORETA solutions were then computed in two fixed time segments (T1: 140-200 ms, T2: 220-290 ms) individually for 5 of the participants who participated in experiment 1B. After establishing the position of each source in Talairach coordinates, anatomical label were obtained with the help of the Talairach Daemon client (Lancaster et al. 2000).

## **Results**

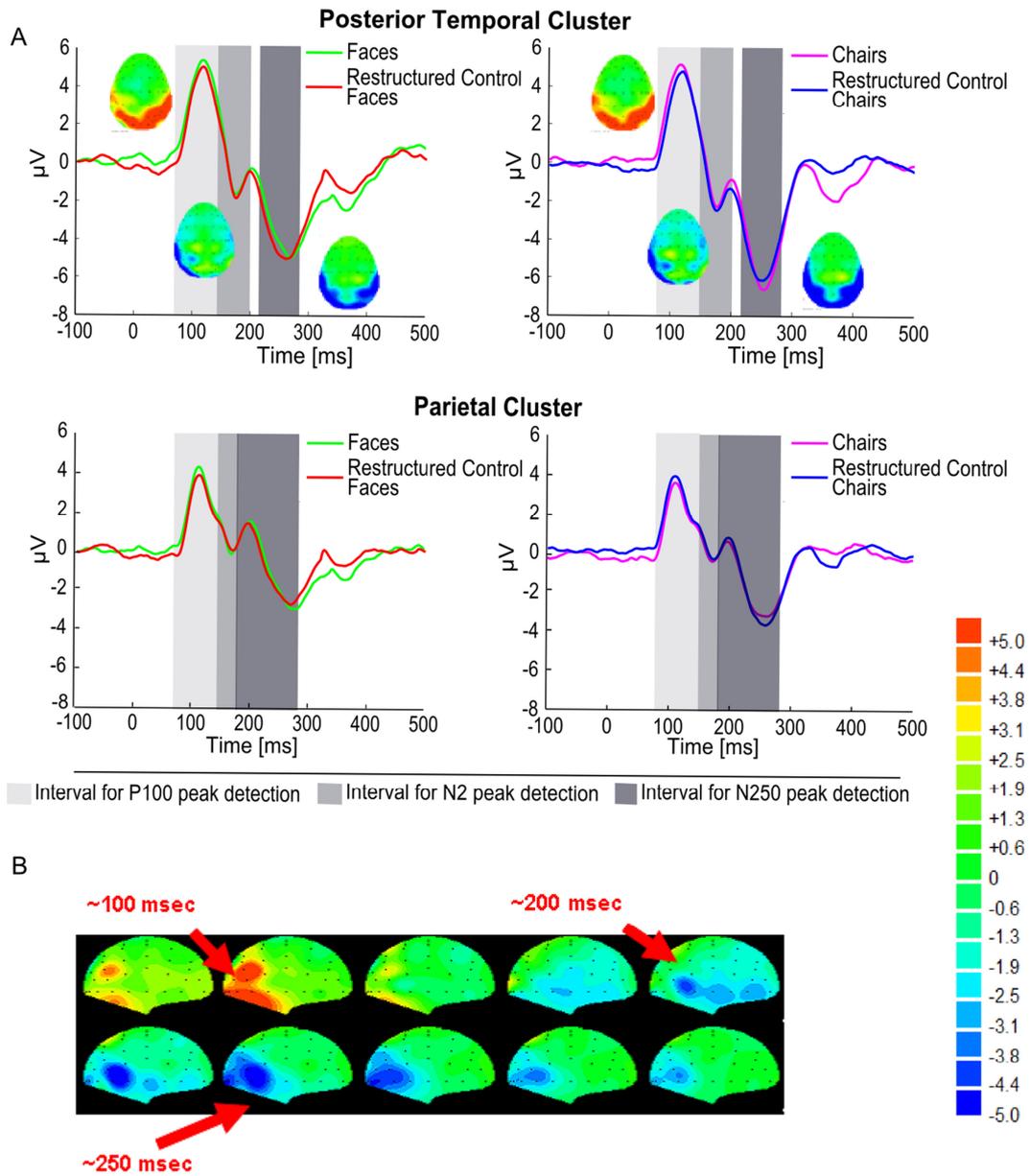
### **Experiment 1A: ERP components elicited by SFM categorical stimuli and matched control stimuli**

Experiment 1 was designed to identify category-related ERP responses elicited by meaningful SFM stimuli of 860 ms duration while participants performed an object identification task (see Materials and Methods). Categories tested were SFM faces and chairs as well as their restructured counterparts (2 categories x 2 controls). ERP results revealed a P100-N2-N250 ERP complex in all 4 conditions (Fig. 2A). In the statistical analysis, separate repeated measures ANOVAs were computed for both amplitude and latency of each of these components. Factors included in the model

were object category (face, chair), stimulus type (intact, restructured) and regional cluster (occipital, parietal, posterior temporal).

For the P100 amplitude, there were no significant main effects of the factors category and stimulus type ( $p > .05$ ) but a significant main effect of cluster ( $F_{(2,14)} = 7.319$ ,  $p = .007$ ), showing that the peaks at the occipital and the parietal clusters were larger than at the posterior temporal cluster. Similarly, the P100 latency analysis revealed no significant main effects of category and stimulus type ( $p > .05$ ) and also no significant main effect of cluster ( $p > .05$ ). ANOVA on the N2 amplitude and latency revealed no significant main effects or interactions ( $p > .05$ ) for any of the factors. For the N250, analysis of peak amplitudes showed no significant main effects (category ( $F_{(1,7)} = 1.553$ ,  $p = .248$ ; stimulus type ( $F_{(1,7)} = 2.040$ ,  $p = .191$ ; cluster ( $F_{(2,14)} = 2.902$ ,  $p = .084$ ) and no significant interactions ( $p > .05$ ). In contrast, analysis of the N250 latency revealed a significant main effect of object category ( $F_{(1,7)} = 7.460$ ,  $p = .026$ ) with face-related stimuli eliciting a later N250 than chair-related stimuli. No other main effects or interaction for the N250 latency were significant ( $p > .05$ ). Voltage maps over time in all conditions showed that early occipital activation at around 100 ms was followed by later (posterior) temporal activation (200 – 300) (Fig. 2B).

In summary, the results show that restructuring the SFM stimuli into meaningless objects, i.e. disrupting the object identity information while leaving visual features intact, did not result in a significant modulation of the ERP amplitude, neither at the N2, nor at the N250. In addition, although there was a significantly longer N250 peak latency for SFM face-related stimuli compared to SFM chair-related stimuli, this increase did not differ between meaningful categorical stimuli and meaningless control stimuli. Thus, neither the amplitude nor the latency of ERP components in this experiment provided information specific for meaningful object categories.



**Figure 2. SFM-evoked ERP signals (Exp. 1A).** **A** The P1-N2-N250 pattern evoked by SFM stimuli of 860 ms duration is displayed for the posterior temporal and the parietal cluster. Voltage maps at corresponding time points are displayed for SFM face and chair elicited ERP signals. Group averaged ERPs showed no differences between SFM objects and associated restructured control stimuli at the N250 peak. The grey shaded areas indicate the time windows used for peak detection. **B** Voltage maps of the time points of interest for SFM faces. Activity in the occipital cortex is followed by more temporal (negative) activation at later time points (200-300 ms).

### **Experiment 1B: ERP components elicited by SFM categorical stimuli and matched control stimuli at brief stimulus durations**

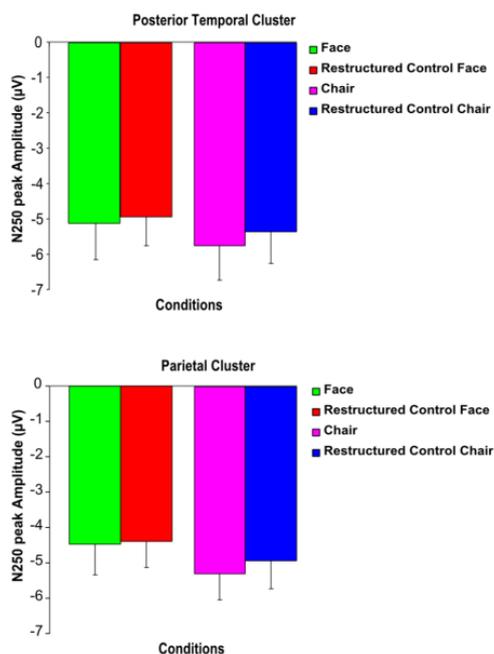
Experiment 1A revealed that the difference between meaningful categorical stimuli and meaningless control stimuli matched on visual features was not reflected in the measured ERP components. For the N2, and especially the N250, it is possible that the lack of categorical object-specificity was due to the long stimulus duration and the object identification task, which may have permitted cognitive factors such as imagery to influence the ERPs elicited by the restructured stimuli. Imagery has been found to have a content-specific effect on the ERP within the first 200 ms of stimulus processing (Farah et al. 1988). Furthermore, in the case of face processing it has been suggested that the early perceptual stage (e.g. the N170) may be penetrable by top-down effects due to the activation of face representations within the face recognition system (e.g. by showing face ‘primes’) (Jemel et al. 2003; Righart et al. 2011). This raises the question whether the face-name association task in experiment 1 of the present study may have influenced the ERP results. Consistent with this idea is also that top-down effects that may be associated with task performance (leading to enhancement of neural activity) can be seen about 150 ms after stimulus onset in static stimuli (Hopf et al. 2002). Hence, we shortened the SFM stimuli duration to 160 ms and applied a simple categorization task that limited the possibility for meaningful object-related cognitive processes to affect the processing of the meaningless restructured control stimuli (for details see Materials and Methods).

Psychophysical test data showed that participants needed SFM stimulus exposures of at least 100 ms to identify object categories emerging from the moving dot pattern (< 10% error rates at duration 100 ms, results not shown) indicating that the 160 ms durations used in the EEG measurements of experiment 1B were sufficiently long to ensure object categorization. The ERP results indicated that despite the shorter stimulus duration, the pattern P100-N2-N250 as reported in experiment 1A was replicated in experiment 1B. Analysis at the P100 amplitude and latency revealed no significant main effects of object category, stimulus type, or cluster ( $p > .05$ ). For the N2 amplitude, analysis showed a significant main effect of regional cluster ( $F_{(2,10)} = 4.271, p = .046$ ) with post-hoc pairwise comparisons revealing that the N2 peak was larger in the posterior temporal than in the occipital cluster ( $p = .012$ ). The remaining factors and interactions did not reach significance ( $p > .05$ ). Similarly, the N2 latency analysis revealed no significant main effects or interactions ( $> .05$ ). ANOVA of N250 amplitudes showed a significant main effect of object category ( $F_{(1,5)} = 7.917, p = .037$ ), corresponding to face-related stimuli eliciting smaller peaks than chair-related stimuli. Importantly, there was no significant difference in amplitude between the intact categorical stimuli and associated meaningless control stimuli ( $F_{(1,5)} = 1.774, p = .240$ ). Regional cluster and the interaction terms did not reach significance ( $p > .05$ ). The N250 latency analysis showed a significant main effect of regional cluster ( $F_{(2,10)} = 8.017, p = .008$ ). Pairwise comparisons indicated that the N250 peaked significantly earlier in the parietal than in the posterior temporal cluster ( $p = .039$ ), suggesting that dorsal

stream processing precedes ventral stream activation during the perception of short-lived SFM stimuli. None of the other main effects or interactions were significant ( $p > .05$ ).

To rule out that the absence of a difference between meaningful and matched meaningless control objects was due to an insufficient number of participants included in the analysis for the two stimulus durations separately, N250 peak amplitude values from all 14 participants who participated in experiments 1A and 1B were entered in a joint ANOVA. Within-subject factors were category, stimulus type, and cluster as well as the between-subject factor stimulus duration. Results revealed a significant main effect of category ( $F_{(1,13)} = 5.171, p = .041$ ), no significant main effect of cluster ( $F_{(2,26)} = 2.698, p = .086$ ) and most importantly no significant main effect of stimulus type ( $F_{(1,13)} = 2.913, p = .112$ ), hence confirming results from the separate analyses of experiments 1A and B (Fig. 3). It could be argued that increasing the number of participants beyond 14 would have revealed a significant effect. If true, our data suggest this effect would be small in size, at best. The between-subject factor duration and the interactions did not reach significance ( $p > .05$ ).

In sum, the data show that by increasing the contribution of bottom-up processing more consistent differences in ERP amplitude between stimuli related to the chair category and stimuli related to the face category can be found at the level of the N250 component. However, we confirmed that the N2 and the N250 do not show an amplitude difference between intact SFM object stimuli and their restructured counterparts showing matched, meaningless objects. The data therefore suggest the possibility that lower-level visual features have a stronger influence on the amplitude of these ERP signals than the categorical aspect of object stimuli.



**Figure 3. Visual features contribute to the amplitude of SFM-evoked ERP signals (Exp. 1A & B).** Bar graphs depicting the mean N250 amplitudes for the SFM object and restructured object conditions of 14 participants pooled for the two durations used in experiment 1 and 2 (160 ms and 860 ms). The results demonstrate that there is no significant difference in the N250 amplitude between meaningful categorical objects and corresponding meaningless control objects in the posterior temporal and the parietal clusters. Error bars indicate the SEM.

### **Experiment 2A: Parametric variation of visual features in categorical stimuli: Psychophysics**

In this experiment, we assessed the impact of visual feature manipulations induced by a reduction of depth range in SFM stimuli (160 ms) on behavioral task performance (Fig. 4A). Overall error rates were low (< 5%) indicating that participants were able to distinguish well between SFM stimulus categories (faces versus chairs). A repeated measures ANOVA testing the effect of the factors depth range (10%, 30%, 90%) and category (face, chair) showed a significant main effect of depth range ( $F_{(2,40)} = 10.354$ ,  $p < .001$ ) with post-hoc comparisons indicating that lowest depth stimuli resulted in higher error rates than highest depth stimuli ( $p = .001$ ). In addition, there was a significant main effect of category ( $F_{(1,20)} = 19.799$ ,  $p < .001$ ) showing that participants committed less errors for the chairs than for the faces. Finally, we found a significant interaction between depth range and category ( $F_{(2,40)} = 10.379$ ,  $p = .001$ ). Post-hoc comparisons confirmed a strong effect of depth range for faces ( $p < .001$ ) but no such effect for chairs ( $p = .956$ ). Thus, these psychophysical data show that while depth range manipulations of SFM stimuli affect performance for face stimuli, they had no effect on chair stimuli.

### **Experiment 2B: Parametric variation of visual features in SFM categorical stimuli: Effects on ERP amplitude**

Experiments 1A and B have shown that the ERP components elicited by SFM-defined categorical stimuli (faces and chairs) were primarily driven by complex visual features that were preserved in the control stimuli, but in themselves did not produce meaningful categorical object perception. To more explicitly reveal the contribution of visual features to these ERP responses, we tested the effect of a parametric variation of depth range in SFM object stimuli (160 ms) on the amplitude of their ERP responses. This parametric variation was achieved by decreasing 3D stimulus depth (90%, 60%, 30%, 10%), thus reducing the relative contribution of complex 3D surface curvature cues and increasing the contribution of simpler 2D cues to object perception. The bar graphs in Figure 4B show the N250 peak amplitudes obtained in the different experimental conditions. A repeated measures ANOVA was performed with the factors depth range (4 levels), category (face, chair) and regional cluster (occipital, parietal, posterior temporal). Since the N2 was small or difficult to detect in some of the conditions, analysis was restricted to the early P100 component and the N250 waveform which was the larger, more consistent ERP peak and thus became the focus of our interest.

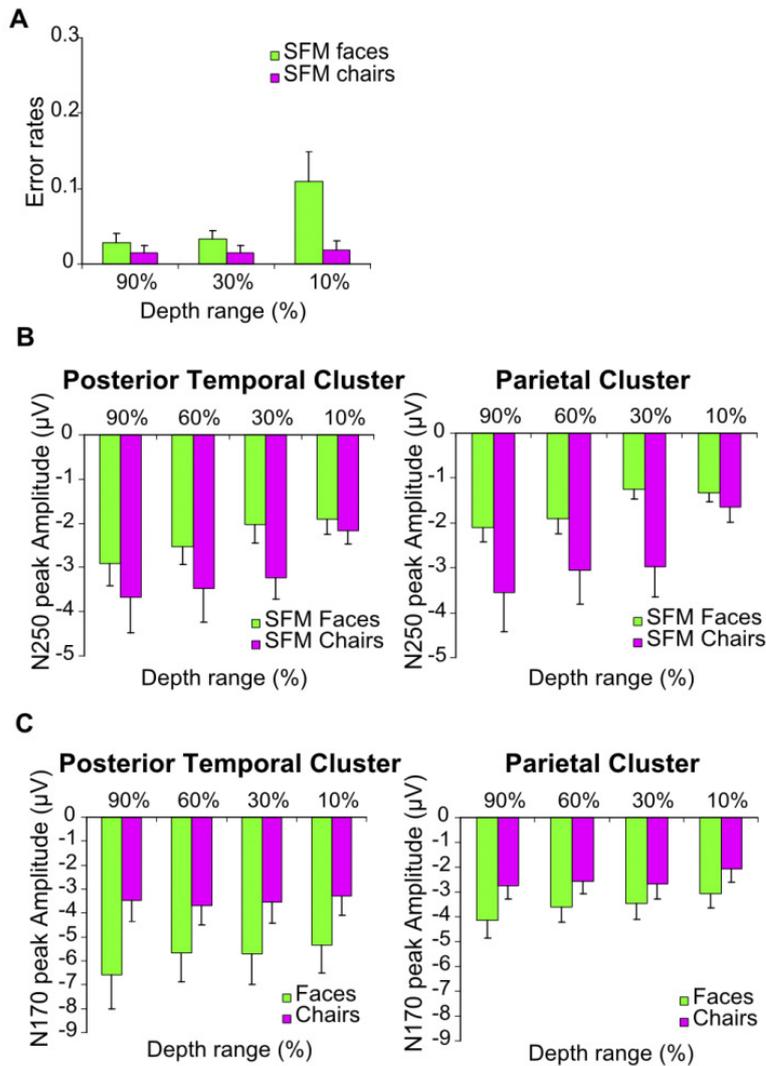
Results for the P100 amplitude showed no significant main effect of depth range, category or regional cluster ( $p > .05$ ) and no significant interaction effects ( $p > .05$ ). For the N250, the more depth was preserved in the stimulus, the larger the negativity was of the peak (see Fig. 4B). This was confirmed by the significant main effect of depth range for the N250 peak ( $F_{(3,18)} = 5.015$ ,  $p = .011$ ) while neither category and cluster nor the interaction terms revealed significant main effects (all  $p$ 's  $> .05$ ). The N250 peak was hence significantly modulated by depth for both

faces and chairs, whereas the behavioural data indicate an effect of depth range on categorization performance only for faces. Thus, the N250 amplitude varied with the depth/complexity of object-related visual features, in a way that was not directly related to categorical object perception or categorization performance.

### **Experiment 2C: Parametric variation of visual features in luminance-defined categorical stimuli: Effects on ERP amplitude**

The previous experiments indicated that variation in the SFM-related N250 ERP peak is determined to a great extent by complex visual feature information that is related to objects, but not by categorical object perception as such. Similar findings have also been obtained by showing sensitivity of the N170 to certain stimulus properties such as spatial frequency (Goffaux et al. 2003; Fievaris et al. 2008) hence contradicting the claim of a direct specific link between the amplitude of ERPs and perception of object categories per se. However, here the psychophysical results suggest that the SFM depth range modulation influenced face categorization (error rates being higher for smaller depth ranges). The possibility hence remains that for the SFM face stimuli, the observed modulation of the N250 peak was due at least in part to the parametric destruction of categorical object perception instead of being attributable to reductions in the complexity and strength of visual object-related features. To test this possibility, we replicated the feature manipulation using luminance-defined faces and chairs (Figure 1B) (500 ms, details in Methods) for which the depth modulation reduced the complexity of visual features but did not have an effect on object recognition ( $p > .05$ ) (error rates  $< 1\%$ , results not shown). The ERP peaks of interest in this case were the P100 and the N170 (Figure 4C). Our repeated measures ANOVA for the P100 amplitude revealed no significant main effects of category or depth range ( $p > .05$ ) but a significant main effect of cluster as well as an interaction between cluster and category ( $F_{(2,8)} = 4.824, p = .042$  and  $F_{(2,8)} = 4.504, p = .024$ , respectively). Separate ANOVAs per category revealed a significant main effect of cluster for the faces ( $F_{(2,8)} = 6.086, p = .025$ ) with the peak being larger in the occipital than in the posterior temporal cluster, but no such effect for the chairs ( $F_{(2,8)} = 3.398, p = .085$ ). Analysis of the N170 peak amplitude revealed significant main effects of depth range ( $F_{(3,12)} = 4.018, p = .034$ ) and cluster ( $F_{(2,8)} = 16.189, p = .002$ ) while the factor category missed statistical significance ( $p > .05$ ). Additionally, there was a significant interaction between category and cluster ( $F_{(2,8)} = 7.762, p = .034$ ). A separate analysis only for face stimuli revealed a significant main effect of depth range ( $F_{(3,12)} = 4.217, p = .030$ ), showing larger peak amplitudes for larger depth ranges (Fig. 4C) as well as a significant effect of regional cluster ( $F_{(2,8)} = 14.419, p = .002$ ), corresponding to larger peak amplitudes in the posterior temporal than in the parietal cluster. The interaction between depth range and cluster was not significant ( $p > .05$ ). The same analysis at the N170 for chairs only revealed a significant effect of regional cluster ( $F_{(2,8)} = 13.637, p = .003$ ) but no significant effect of depth ( $F_{(3,12)} = .224, p = .878$ ) or interaction effect ( $p > .05$ ).

In sum, we find that a reduction of stimulus depth range modulates the N170 response to faces without compromising object categorization as measured psychophysically. Additionally and in agreement with the current literature our results also indicated higher amplitudes of the N170 to static face compared to static chair stimuli in the posterior temporal cluster.



**Figure 4. Depth range modulation of ERP signals (Exp. 2).** **A** Behavioral data indicate that a reduction in depth range of the SFM stimuli decreased recognition performance of faces but not of chairs. Bar graphs indicate the error rate and error bars indicate the SEM. **B** The N250 is modulated by SFM face and chair depth range in both posterior temporal and parietal clusters. **C** The N170 to static faces but not to static chairs is modulated by depth range in posterior temporal and parietal clusters. N170 amplitude in the posterior temporal cluster is larger in response to faces compared to chairs. Bar graphs show the mean amplitude of the peaks and error bars the SEM.

### Source localization

Source localization was performed for the SFM face and chair conditions in experiment 1B (with short duration SFM stimuli), separately for 5 of the 6

participants in this experiment. In the early time window (140–200 ms), we found in both conditions activity in dorsal regions, in the proximity of putative motion areas MT and V3 as well as activation in left superior parietal and right fusiform areas. In the later time window (200–290 ms) in addition to dorsal sources, the SFM object stimuli activated the right fusiform gyrus as well as regions around the right STS and the right lateral occipital cortex, consistent with dorsal-ventral integration in recognition of motion-defined objects (Farivar 2009; Farivar et al. 2009). In all participants and for both conditions we observed a right hemispheric dominance. In sum, source localization of SFM stimuli yielded results suggesting a shift from sources for visual feature analysis in early ERPs, to more high-level sources in later ERPs.

## **Discussion**

In the present study, we investigated the perceptual correlates of ERP signals elicited by meaningful SFM- and luminance-defined objects. Since in SFM perception local motion cues define object categories, we asked whether category-specific responses could be isolated from visual feature-related response representations in the ERP signal. Interestingly, the SFM defined face and chair stimuli induced N2 and N250 peaks that showed little specificity for the categorical aspects of the stimuli. Instead, ERP peak amplitudes were highly sensitive to visual feature properties and were strongly modulated by visual stimulus feature manipulations.

These findings raise several questions: First, given the high impact of visual features on object-related ERP peaks in the present experiment, what general conclusions can we draw if an ERP peak varies its amplitude with object category? Second, can we interpret the N250 in response to SFM objects as a delayed N170? Third, how can we solve the inverse problem of relating an event-related potential with a stimulus that is characterized by a large set of dimensions on a variety of scales of complexity?

### **The relative contribution of visual features and categorical aspects to object-related ERPs**

Numerous studies have related ERP responses to category- (face-) specific processing (e.g. Bentin et al. 1996; Eimer and McCarthy 1999; Rossion and Gauthier 2002). For this hypothesis to be supported, changes in the stimulus that change categorical perception should strongly modulate ERP amplitude, and changes in visual features that do not affect categorical perception should not or less strongly do so. However, in our data, we found the opposite, thereby raising the general question about contributions of low- to intermediate-level visual features in determining object-related ERPs. Our study does not deny the relative relevance of categorical processing but emphasizes the contribution of intermediate level features to ERP responses. The data we obtained pertain both to the N170

traditionally linked with categorical perception in static stimuli, and to the N250 peak we have observed for SFM objects.

Prior studies using static face stimuli have provided important insights into the extent to which ERPs can be linked to categorical perception by showing that removing outer face contours (Zion-Golumbic and Bentin 2007), presenting features in isolation (Bentin et al. 1996) or changing the type of face representation (schematic, photographic etc) (Latinus and Taylor 2006) resulted in N170 signal modulations, whereas changing the configuration of inner face components (ICs) without changing the contour of the face did not (e.g.(Zion-Golumbic and Bentin 2007)). These results indicate that changes of visual stimulus features result in N170 peak modulations while manipulations leaving these features and surface structures intact fail to do so. Moreover, the N170 to static faces shows less pronounced peak amplitudes for Mooney faces (made of simple object cues) compared to photographic and schematic faces (Latinus and Taylor 2006) hence supporting this claim and indicating a role of visual feature complexity in determining ERP responses. This view is in line with results from our parametric feature manipulation experiment with static stimuli in which a decrease in depth range in face stimuli led to a significant decrease in N170 amplitude. Our findings indicate that a simple manipulation of visual depth cues in the image can modify N170 amplitudes in response to faces without affecting categorical perception, making an unambiguous interpretation of this ERP difficult. These findings complement similar results in previous studies pointing to a dependence of N170 category effects on parameters such as spatial frequency (Flevaris et al. 2008) and inter stimulus perceptual variance (Thierry et al. 2007). It should be acknowledged that our findings on the N170 are based on a relatively small pool of participants. Nevertheless, these results contribute to converging evidence from multiple experiments suggesting that a link between the amplitude of object related ERPs and categorical object perception is not straightforward.

Additionally, in the present study we observed an N250 component in response to SFM objects and investigated also for this component whether it can be linked unambiguously with categorical object perception. Here, we found again that a manipulation of depth range, causing a decrease in strength and complexity of several visual features but retaining the 'categorical objectness' of the stimuli led to a strong decrease in amplitude for the SFM-related N250. This effect was unlikely to be explained by changes in categorical perception as shown by psychophysical control data. In addition, the N250 ERP amplitudes were not significantly modulated by stimulus manipulations that destroyed categorical perception while keeping other lower level visual features constant.

The converging evidence suggests that from late ERP amplitudes (N170, N250) it is difficult to derive object-specificity since it depends on the values of visual parameters that within certain ranges do not influence object categorization but do modulate ERP amplitude. Other indicators such as the face-inversion effect (Rossion et al. 1999, 2000) of the N170 latency might therefore be a more reliable correlate of face-specific processing since in this manipulation most low-level visual properties remain intact while our face-processing expertise is destroyed.

## **The N250: A delayed N170?**

Among the numerous ERP studies that have aimed to investigate the time course of category-related effects in visual perception, to our knowledge, none have used meaningful and complex structure-from-motion 3D objects. With static stimuli, face-specific effects have mostly been reported at 170 ms (Rossion et al. 2000; Itier 2004; Itier and Taylor 2004) while in the present study using SFM-defined stimuli, no such face specificity was found in this time window. Our psychophysical experiments indicated that participants needed stimulus durations of at least 100 ms to identify object categories emerging from the moving dot pattern. This result is in line with a study comparing ERPs and behavioral responses to simple objects (sphere and cylinder) defined by luminance or motion, indicating a delay of 80 msec both in MEG/EEG responses and in reaction times when additional motion information was processed (Jiang et al. 2008). This suggests that the lack of a face-specific effect at 170 ms with SFM stimuli in the present study may be due to higher processing load which caused a shift of the component to 250 ms (N250). To test whether the N250 is in fact an N170 analogue, future studies could investigate face-inversion effects for SFM faces at this peak. Additionally, studies aiming to more accurately study the timing of SFM-related components could lock the ERP signals to button presses signaling categorization instead of stimulus onset. However, a latency shift of the N170 remains a matter of speculation and is not straightforward since the N250 was not face-specific. This lack of face specificity, however, might be due to a dominant contribution of lower-level and dynamic visual features to the ERP amplitudes, masking a relatively weak contribution of category-specific information, which in the context of our study would support the interpretation of the N250 as a delayed N170.

Irrespective of the precise interpretation of the N250, our study indicates that N250 amplitudes in response to SFM objects cannot easily be linked with object categorization, given their strong dependency on visual cues in the image that however do not affect categorization performance.

## **Inverse problem**

Categorical stimuli constitute a challenge for the analysis of their ERP responses since they are characterized by a large set of stimulus dimensions. Object category, cue configurations, cue complexity/depth, dynamic features as well as high-level factors may each contribute to the observed signal posing an interesting inverse problem. Below, we briefly discuss contributions from a subset of relevant factors.

The N2 we report here has been described in previous studies investigating the ERP responses to motion onset perception (Kuba and Kubová 1992; Schlykova et al. 1993; Bach and Ullrich 1994) and is believed to originate from MT (Schlykova et al. 1993). The human motion complex hMT+ (Zeki et al. 1991) (Tootell et al. 1995) in its medial superior temporal location has repeatedly been shown to play a role in SFM perception (Orban et al. 1999; Paradis et al. 2000; Kriegeskorte et al. 2003). However, also at later latencies motion-related effects have been reported.

An additional negative peak at 240 ms over occipitotemporal regions which, in contrast to the N2 is believed to reflect the higher processing of motion stimuli (Hirai et al. 2003), may also be contributing to the N250 peak amplitude in the present study.

Effects of depth cues on the ERP signal have previously been reported for static stimuli indicating a negative potential around 170 ms present on occipital sites which was consistently enhanced to depth cues (Jeffreys 1996). Furthermore, fMRI revealed that area hMT+ is part of a network with right hemispheric dominance including a lateral occipital region, five sites along the intraparietal sulcus (IPS), and two ventral occipital regions that is involved in extracting depth from motion (Orban et al. 1999). The depth modulations of the ERP peaks in the present study are likely to be driven by computations taking place in this entire network.

Although there seems to be a strong influence of feature-based properties on ERP signals, it is not the case that ERPs cannot pick up category-related effects. Paradigms using stimuli in which category percepts pop out once certain structures/configurations are detected without any change in the stimulus' physical features provide strong evidence for more isolated contributions of categorical object classification to ERP signals (Bentin and Golland 2002; Pegna et al. 2004; George et al. 2005).

Hence, many visual features in an object stimulus could influence the amplitude of ERPs in addition to its categorical aspect. Although the data we have obtained were collected in a set of experiments with relatively small numbers of participants in the individual experiments, the converging results from the data as a whole (on 27 participants) support this idea. They indicate that the amplitude of late ERP components that are elicited by object stimuli is difficult to link unambiguously with categorical object perception. This is shown both by the lack of ERP amplitude modulation when specific categorical perception is destroyed but low-to-mid-level visual features are retained and by the strong and significant ERP amplitude modulations in stimuli that retain their categorical aspect but that are changed on the strength of low-to-mid-level visual features. Thus, our data indicate that the contributions of categorical and visual processing are difficult to distinguish, pointing to a severe inverse problem in interpreting those ERP components. While measuring changes in categorical perception in physically unchanged stimuli may be one approach to tackle this problem, new ways of analyzing ERP data that no longer rely on trial averaging may provide another. Specifically, Bayesian approaches and classification algorithms like SVMs (Makeig et al. 2002; Sparacino et al. 2002; Mueller et al. 2010) might be promising tools to overcome the inverse problem faced by many studies in the ERP field.

## References

- Allison T, Puce A, Spencer DD, McCarthy G. 1999. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex*. 9:415–430.
- Bach M, Ullrich D. 1994. Motion adaptation governs the shape of motion-evoked cortical potentials. *Vision Res*. 34:1541–1547.
- Bentin S, Allison T, Puce A, Perez E, McCarthy G. 1996. Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*. 8:551–565.
- Bentin S, Golland Y. 2002. Meaningful processing of meaningless stimuli: the influence of perceptual experience on early visual processing of faces. *Cognition*. 86:B1–14.
- Caharel S, Courtay N, Bernard C, Lalonde R, Rebaï M. 2005. Familiarity and emotional expression influence an early stage of face processing: an electrophysiological study. *Brain Cogn*. 59:96–100.
- Caharel S, Poiroux S, Bernard C, Thibaut F, Lalonde R, Rebai M. 2002. ERPs associated with familiarity and degree of familiarity during face recognition. *Int J Neurosci*. 112:1499–1512.
- Eimer M. 2000. The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport*. 11:2319–2324.
- Eimer M, McCarthy R a. 1999. Prosopagnosia and structural encoding of faces: evidence from event-related potentials. *Neuroreport*. 10:255–259.
- Farah MJ, Péronnet F, Gonon M a, Giard MH. 1988. Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *J Exp Psychol Gen*. 117:248–257.
- Farivar R. 2009. Dorsal-ventral integration in object recognition. *Brain Res Rev*. 61:144–153.
- Farivar R, Blanke O, Chaudhuri A. 2009. Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *J Neurosci*. 29:5336–5342.
- Flevaris AV, Robertson LC, Bentin S. 2008. Using spatial frequency scales for processing face features and face configuration: an ERP analysis. *Brain Res*. 1194:100–109.
- George N, Jemel B, Fiori N, Chaby L, Renault B. 2005. Electrophysiological correlates of facial decision: insights from upright and upside-down Mooney-face perception. *Brain Res Cogn Brain Res*. 24:663–673.
- Goffaux V, Gauthier I, Rossion B. 2003. Spatial scale contribution to early visual differences between face and object processing. *Brain Res Cogn Brain Res*. 16:416–424.
- Haushofer J, Baker CI, Livingstone MS, Kanwisher N. 2008. Privileged coding of convex shapes in human object-selective cortex. *J Neurophysiol*. 100:753–762.

- Herrmann MJ, Ehlis A-C, Ellgring H, Fallgatter a J. 2005. Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *J Neural Transm.* 112:1073–1081.
- Hirai M, Fukushima H, Hiraki K. 2003. An event-related potentials study of biological motion perception in humans. *Neurosci Lett.* 344:41–44.
- Hopf J, Vogel E, Woodman G, Heinze H-J, Luck SJ. 2002. Localizing visual discrimination processes in time and space. *J Neurophysiol.* 88:2088–2095.
- Itier R, Taylor M. 2004. Source analysis of the N170 to faces and objects. *Neuroreport.* 15:1–5.
- Itier RJ. 2004. N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs. *Cereb Cortex.* 14:132–142.
- Itier RJ, Taylor MJ. 2002. Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuroimage.* 15:353–372.
- Jacques C, Rossion B. 2007. Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *Neuroimage.* 36:863–876.
- Jeffreys D a. 1996. Simple methods of identifying the independently generated components of scalp-recorded responses evoked by stationary patterns. *Exp Brain Res.* 111:100–112.
- Jemel B, Pisani M, Calabria M, Crommelinck M, Bruyer R. 2003. Is the N170 for faces cognitively penetrable? Evidence from repetition priming of Mooney faces of familiar and unfamiliar persons. *Brain Res Cogn Brain Res.* 17:431–446.
- Jiang Y, Boehler CN, Nönnig N, Düzel E, Hopf J-M, Heinze H-J, Schoenfeld MA. 2008. Binding 3-D object perception in the human visual cortex. *J Cogn Neurosci.* 20:553–562.
- Kayaert G, Biederman I, Vogels R. 2005. Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cereb Cortex.* 15:1308–1321.
- Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. *J Neurosci.* 23:1451–1463.
- Kuba M, Kubová Z. 1992. Visual evoked potentials specific for motion onset. *Doc Ophthalmol.* 80:83–89.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten S a, Fox PT. 2000. Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp.* 10:120–131.
- Latinus M, Taylor MJ. 2006. Face processing stages: impact of difficulty and the separation of effects. *Brain Res.* 1123:179–187.
- Liu CH, Collin C a, Farivar R, Chaudhuri A. 2005. Recognizing faces defined by texture gradients. *Percept Psychophys.* 67:158–167.

- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ. 2002. Dynamic brain sources of visual evoked responses. *Science*. 295:690–694.
- Motter B. 1994. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J Neurosci*. 14.
- Mueller A, Candrian G, Kropotov JD, Ponomarev V a, Baschera G-M. 2010. Classification of ADHD patients on the basis of independent ERP components using a machine learning system. *Nonlinear Biomed Phys*. 4:1–12.
- Orban G a, Sunaert S, Todd JT, Van Hecke P, Marchal G. 1999. Human cortical regions involved in extracting depth from motion. *Neuron*. 24:929–940.
- Paradis a L, Cornilleau-Pérès V, Droulez J, Van De Moortele PF, Lobel E, Berthoz A, Le Bihan D, Poline JB. 2000. Visual perception of motion and 3-D structure from motion: an fMRI study. *Cereb Cortex*. 10:772–783.
- Pascual-Marqui R, Michel C, Lehmann D. 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiology*. 18:49–65.
- Pascual-Marqui RD. 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*. 24:5–12.
- Pegna AJ, Khateb A, Michel CM, Landis T. 2004. Visual recognition of faces, objects, and words using degraded stimuli: where and when it occurs. *Hum Brain Mapp*. 22:300–311.
- Righart R, Burra N, Vuilleumier P. 2011. Face perception in the mind's eye. *Brain Topogr*. 24:9–18.
- Rossion B, Delvenne JF, Debatisse D, Goffaux V, Bruyer R, Crommelinck M, Guérit JM. 1999. Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol Psychol*. 50:173–189.
- Rossion B, Gauthier I. 2002. How does the brain process upright and inverted faces? *Behav Cogn Neurosci Rev*. 1:63–75.
- Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, Crommelinck M. 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*. 11:69–74.
- Rossion B, Jacques C. 2011. The N170: understanding the time-course of face perception in the human brain. In: Luck S, Kappenman E, editors. *The Oxford handbook of ERP components*. Oxford, Uk: Oxford University press.
- Rousselet G, Husk J, PJ B, Sekuler A. 2008. Time course and robustness of ERP object and face differences. *J Vis*. 8:1–18.
- Rousselet G a, Pernet CR, Bennett PJ, Sekuler AB. 2008. Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neurosci*. 9:98.

- Schlykova L, van Dijk BW, Ehrenstein WH. 1993. Motion-onset visual-evoked potentials as a function of retinal eccentricity in man. *Brain Res Cogn Brain Res*. 1:169–174.
- Schweinberger SR, Huddy V, Burton a. M. 2004. N250r: a face-selective brain response to stimulus repetitions. *NeuroReport*. 15:1501–1505.
- Sparacino G, Milani S, Arslan E, Cobelli C. 2002. A Bayesian approach to estimate evoked potentials. *Comput Methods and Programs Biomed*. 68:233–248.
- Thierry G, Martin CD, Downing P, Pegna AJ. 2007. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci*. 10:505–511.
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J of Neurosci*. 15:3215–3230.
- Troje N, Bühlhoff H. 1996. Face recognition under varying poses: The role of texture and shape. *Vision Res*. 36:1761–1771.
- VanRullen R, Thorpe SJ. 2001. The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci*. 13:454–461.
- Vizioli L, Foreman K, Rousselet G, Caldara R. 2010. Inverting faces elicits sensitivity to race on the N170 component: A cross-cultural study. *J Vis*. 10:1–23.
- Vlamings P, Goffaux V, Kemner C. 2009. Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? *J Vis*. 9:1–13.
- Zeki S, Watson J, Lueck C. 1991. A direct demonstration of functional specialization in human visual cortex. *J Neurosci*. 17:641–649.
- Zion-Golumbic E, Bentin S. 2007. Dissociated neural mechanisms for face detection and configural encoding: evidence from N170 and induced gamma-band oscillation effects. *Cereb Cortex*. 17:1741–1749.

## Impaired Processing of 3D Motion-Defined Faces in MCI and Healthy Ageing: An fMRI Study

---

*Corresponding publication:* Graewe B, Lemos R, Ferreira C, Santana I, Farivar R, De Weerd P, Castelo-Branco M. (2012) Impaired Processing of 3D Motion-Defined Faces in Mild Cognitive Impairment and Healthy Aging: An fMRI Study. *Cereb Cortex*. doi: 10.10.

## **Abstract**

Mild cognitive impairment (MCI), which shows high risk for conversion to Alzheimer's disease (AD), is accompanied by progressive visual deteriorations that so far are poorly understood. Here, we compared dorsal and ventral visual stream functional magnetic resonance imaging (fMRI) activity among amnesic MCI, healthy elderly and young participants during structure-from-motion (SFM) face categorization performance. Task performance varied with stimulus depth and duration levels and differences among groups were highly correlated with face-related fMRI activation patterns. Young participants showed larger activation to faces than scrambled faces (face sensitivity) in right FFA/OFA whereas in elderly, this difference was reduced. Surprisingly, in MCI, scrambled faces elicited larger activation in right FFA/OFA than faces. The latter observation may be related to the additional finding of elevated depth sensitivity in left FFA/OFA of MCI, suggesting that an increased representation of low-level stimulus aspects may impair face perception in MCI. Discriminant function analysis using face and depth sensitivity indices in FFA/OFA classified MCI and healthy elderly with 88.2 % accuracy, marking a fundamental distinction between groups. Potentially related findings include altered activation patterns in dorsal-ventral stream integration regions and attention-related networks of MCI patients. Our results highlight aberrant visual and additional potentially compensatory processes that identify dispositions of (preclinical) AD.

## Introduction

Mild Cognitive Impairment (MCI) represents a boundary cognitive stage between normal aging and dementia (Petersen et al. 1999). Among the several subtypes that have been identified, in particular amnesic MCI patients show an accelerated conversion to AD (Alzheimer's disease) with a yearly transition rate of 10-15% (Petersen et al. 2001). Despite there being no cure or prevention for AD at present, medication for modifying or slowing down the progression of the disease is currently being developed, making biomarkers for early detection of true preclinical AD a major concern.

Apart from memory complaints, AD symptoms also include higher visual dysfunctions. Behavioural studies indicate that the most prominent visual impairments are in the domain of visuospatial function such as stereopsis (Mendez et al. 1996) and (self-) motion perception (Mapstone et al. 2003, 2006; Kavcic et al. 2011) with the latter already being present at the MCI stage. However, in AD, face- (Sauer et al. 2006) and object-recognition impairments (Cronin-Golomb et al. 1995; Mosimann et al. 2004) have also been reported. In the framework of a ventral-dorsal pathway dichotomy (Ungerleider and Mishkin 1982), these findings suggest the dorsal stream as an early and major target of neuropathological changes in the course of the disease, but also potential deficits in ventral stream processing.

These insights have been corroborated by structural neuroimaging studies, which have mainly focused on methods such as voxel-based morphometry (Pennanen et al. 2005; Bozzali et al. 2006), diffusion tensor imaging (DTI) (Cherubini et al. 2010; Gold et al. 2011) and more recently also on cortical thickness measurements (Hampel et al. 2008; Li et al. 2011). Reductions in the cortical grey matter in the mediotemporal and parietal lobe (Baron et al. 2001; Pennanen et al. 2005; Singh et al. 2006) as well as in frontal regions (Baron et al. 2001; Singh et al. 2006) may all contribute to the observed visual deficits in MCI and AD. Apart from deviations in brain *structure*, alterations in brain *function* of MCI have also been addressed by numerous resting state studies (e.g. Fleisher et al. 2009). However, to our knowledge there are very few task-related functional neuroimaging studies. A notable exception are the preliminary data shown in a review article by Yamasaki et al. (2012) suggesting decreased activation in the inferior parietal lobe (IPL) in response to optic flow in MCI in comparison with controls. Furthermore, Bokde et al. (2008) found that MCI patients activated both visual pathways during a location matching task while healthy controls showed selective activation of the dorsal stream. This, along with a study by Bokde et al. (2006) using functional connectivity, confirms the presence of not only dorsal but also ventral visual stream alterations in MCI brain function.

Here, we used concurrent psychophysics and fMRI to characterize functional alterations in both dorsal and ventral visual pathways in the normal ageing and MCI brain. To that goal, we used structure-from-motion (SFM) face stimuli in which motion integration mechanisms in the dorsal stream (Andersen and Bradley 1998) form the basis for object recognition in the ventral stream. These stimuli, apart

from probing dorsal-ventral cross-talk provide an additional window on temporal lobe processes related to face perception in the healthy ageing and MCI. While prior psychophysical studies have shown SFM deficits to be highly correlated with MCI (Lemos et al. 2012) and AD severity (Kim and Park 2010), we here aimed to differentiate both psychophysical performance and associated fMRI brain activation between MCI-afflicted and normal ageing participants. As MCI-related changes in brain activity can be subtle and perception of objects based on motion integration is critically dependent on time, we maximized our ability to differentiate by manipulating task difficulty through variations in stimulus depth and duration. Based on the notion of disturbed integrative function in MCI, we hypothesized an impaired ability to distinguish motion-defined intact from scrambled faces in MCI patients. The contributions to this impairment were investigated in dorsal areas as well as ventral face selective regions for which holistic processing is particularly essential.

## Methods

### Participants

Forty-five right-handed participants (13 young, 17 elderly and 15 MCI) took part in the study. After applying experimental exclusion criteria, i.e. poor or no task performance due to a lack of understanding and excessive movement during the fMRI acquisition (see details below), 12 young participants, 13 elderly participants and 13 amnesic MCI patients were included in the analysis. Young participants were healthy graduate students from the University of Coimbra who volunteered to participate in the experiment. The elderly control group was taken from our database of volunteers or consisted of University or hospital staff as well as their relatives with normal ophthalmological and neurological examination. To exclude cognitive impairment, healthy elderly participants underwent neuropsychological screening (Mini-Mental Status Examination (MMSE) (Folstein et al. 1975)). In the MCI patients the diagnostic investigation additionally included a standard clinical evaluation, a staging assessment, laboratory tests including Apolipoprotein E allele genotyping as well as imaging studies (computed tomography or MRI and single photon emission computed tomography [SPECT]) to exclude comorbidities. MCI patients were recruited from the Neurology Department of the Coimbra University Hospital where diagnosis was achieved through gold standard neurological and neuropsychological assessment following classification criteria for MCI (Petersen 2004, Albert et al. 2011). Only amnesic “*single domain*” MCI patients were included in this study. MCI patients were identified as follows: 1) A subjective complaint of memory decline (reported by the participant or an informant); 2) An objective memory impairment (considered when scores on a standard revised Wechsler Memory Scale (Wechsler 1987) [Immediate and Delayed logical memory] were  $> 1.5$  SDs below age/education adjusted norms); 3) Normal general cognition suggested by normal scores in the MMSE and ADAS-Cog (Portuguese versions)

using the Portuguese cut off scores (Guerreiro et al. 2008a, 2008b); 4) Largely normal daily life activities; 5) Absence of dementia and 6) Exclusion of comorbidities using imaging methods. Exclusion criteria were neurological/psychiatric conditions other than MCI and abnormal ophthalmological conditions. MCI and elderly control participants were matched for chronological age and education level ( $p > .05$ ). Table 1 summarizes the participant demographic data.

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Coimbra. All participants provided verbal and written informed consent before participating in the study. To ensure that participants had understood the task, the fMRI session was preceded by a brief training session outside the scanner. Participants lay supine in the scanner and were given earplugs and padding to maximise comfort and minimize involuntary head movements. A video recording system (RealEye 5721, Avotec Inc.) inside the scanner was used to monitor the elderly and MCI participants ensuring safety and wakefulness. At any time participants had access to a button signalling that they wished to interrupt the session.

**Table 1.** Participant demographic data

	Young participants ( <i>N</i> =12)		Elderly participants ( <i>N</i> =13)		MCI patients ( <i>N</i> =13)	
	<i>Mean</i>	<i>(SD)</i>	<i>Mean</i>	<i>(SD)</i>	<i>Mean</i>	<i>(SD)</i>
<b>Chronological age (years)</b>	26.1	(3.9)	63.6	(6.6)	66.5	(5.8)
<b>Education level (years)</b>	-	-	11.2	(4.9)	8.0	(4.6)
<b>MMSE score</b>	-	-	29.5	(.9)	28.9	(1.4)
<b>Gender (m:f)</b>	4:8		7:6		6:7	
<b>APOE E4 allele carriers</b>	-		-		7	

## Experimental stimuli

The fMRI session consisted of the SFM task and subsequent localizer scans to define regions of interest (ROIs). The SFM stimuli and the face localizer stimuli were presented using Presentation 12.1 software (Neurobehavioral systems) and the motion localizer stimuli were presented using MATLAB (R2009b). Stimuli were presented using an INM PC and were projected with an LCD projector (SilentVision 6011, Avotec Inc, Florida USA) onto a rear-projection Fujitsu Siemens type screen (1024 x 768, refresh rate 60 Hz) located at the head of the participants. The screen was viewed at a distance of 45.5 cm with an angled mirror positioned on the head-coil and subtended 30.3 deg width and 23.1 deg height of visual angle.

### *SFM stimuli*

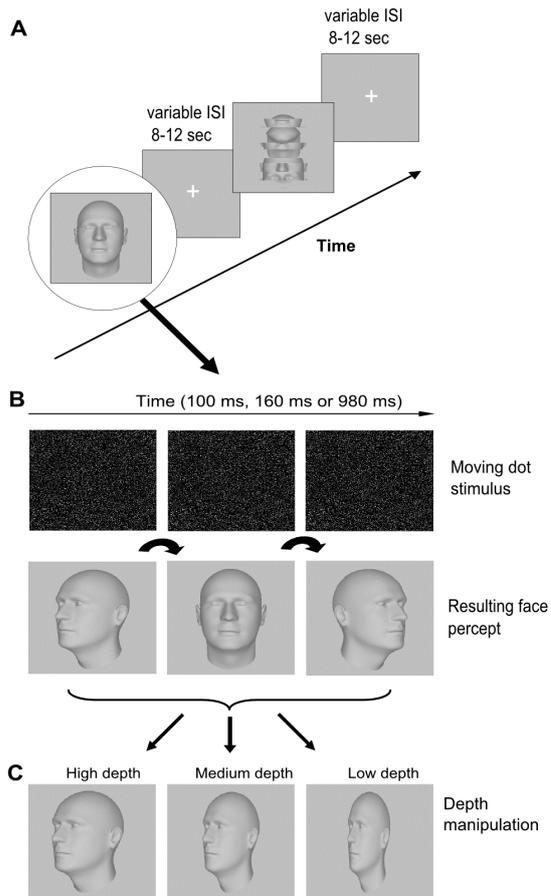
SFM stimuli were videos of motion-defined male faces and scrambled control faces (for details see Farivar et al. 2009) (Fig. 1). Face stimuli were constructed from a

laser-scanned facial surface taken from the Max-Planck Face Database. The surface was rendered with a volumetric texture map that ensures uniform texture density (as described in detail in Liu et al. 2005). Shadows and shading were removed and the face was rendered against a similarly textured random-dot background. Control versions of the face stimuli were constructed by cutting the rendered whole object videos in the horizontal plane into ten blocks and scrambling their positions. This scrambling manipulation resulted in control stimuli that were equated on many of the visual features of the original face videos (including luminance, contrast, texture, spatial frequency, motion, and strength of curvature cues in surfaces) but showed meaningless object entities. Stimuli subtended  $\sim 10.6$  deg of visual angle horizontally and 8.1 deg vertically.

To increase sensitivity to MCI-related deficits, the stimulus variables duration and depth were parametrically manipulated in 3 steps, resulting in an overall  $2(\text{stimulus type}) \times 3(\text{durations}) \times 3(\text{depth levels})$  design. The 3 different duration conditions were  $\sim 980$  ms,  $\sim 160$  ms and  $\sim 100$  ms (Fig. 1B). In each condition, stimuli rotated from a left to right profile view with the extent of rotation being  $-22.5$  deg to  $22.5$  deg,  $-4.3$  deg to  $4.3$  deg, and  $-2.6$  deg to  $2.6$  deg, respectively in long, medium and short duration conditions (frontal view = 0 deg). The global manipulation of the variable 'depth' (Fig. 1C) reduced 3D information across the stimuli resulting in new stimuli with 3 different depth levels (10%, 60% and 90%) parameterized relative to the full anterior-posterior range of the non-transformed 3D stimuli (i.e., 10 % corresponding to a near-flat stimulus). The rotation speed was held constant among the different depth levels. Reducing depth caused a shift from surface curvature cues for face detection to simpler 2D (contour-related) cues. Although the analysis of gender identity was irrelevant for our study, the definition of faces by short-lived motion cues likely did not encourage the detailed luminance and colour analysis of eyes, eyebrows and mouth regions, which have been shown to be the most reliable cues for face gender discrimination in photographic stimuli (Dupuis-Roy et al. 2009).

### ***Localizer stimuli***

Localizer stimuli used to define ROIs for object and face perception were static grayscale images of faces, places (landscapes, skylines), objects (tools, cars, chairs), and scrambled versions of objects ( $9.5 \times 9.5$  deg). Stimuli were presented on a dark background centred at fixation (fixation cross with radius  $0.15^\circ$  in middle of display). MT localizer stimuli ( $12.6 \times 12.6$  deg) were textures that were either static or coherently moving, consisting of arrays of 600 white dots ( $0.3$  deg diameter) on a dark background. Dots coherently moved in one of 4 possible directions (right, left, up, down) at a constant speed ( $5$  deg/s).



**Figure 1. Stimuli and Paradigm** A SFM faces and scrambled faces were presented randomly at 1 of 3 durations and 3 depth levels separated by a variable 8-12 sec ISI B Each SFM stimulus rotated from left to right in one cycle and appeared for 100, 160 or 980 ms. The moving dot pattern provided the only cue to object perception, rendering the object invisible when the rotation/motion stopped. C The depth manipulation resulted in SFM stimuli with 3 depth levels (high, medium and low depth) parameterized in terms of anterior-posterior range. In ‘flattened’ stimuli the distance between the tip of the nose and the back of the head was parametrically reduced. Contrast-defined stimuli in A-C are used merely to illustrate the percepts induced by SFM.

## Behavioural and fMRI data acquisition

Behavioural responses in the scanner were collected via a button box (right and left hand index finger button presses). Participants underwent event-related fMRI in a 3T Siemens Trio scanner (Erlangen, Germany). Functional images were acquired using BOLD contrast echo-planar imaging (EPI, TR=2.5, TE=49, interleaved slice acquisition order, voxel size 2×2×4 mm, 29 slices, no gap, in-plane matrix 128×128, Flip angle 90 deg, field of view [FOV]=256 mm) covering the entire brain. The scanning session included motion- and face localizers (EPI, TR=2, TE=47, interleaved slice acquisition order, voxel size 2.5×2.5×3.5, 25 slices, no gap, in-plane matrix 102×102, Flip angle 90 deg, FOV=256 mm) and a high-resolution T1-weighted anatomical scan (magnetization-prepared-rapid-acquisition-gradient-echo (MP-RAGE) sequence, TR=2.3, TE=2.98, voxel size 1×1×1 mm, 160 slices, in-plane matrix 256×256, Flip angle 9 deg, FOV=256 mm).

## Experimental design and behavioural tasks

### *SFM runs and task*

The SFM experiment was divided into 4 separate sequential runs lasting approximately 6 minutes each. We employed a slow event-related design comprising 18 conditions (stimulus type (intact, scrambled) x duration (short, medium, long) x depth level (high, intermediate, low)) that were presented

randomly within a run. There were 8 trials per condition in total. Participants of all 3 groups performed an SFM face detection task in which they indicated by a button press whether the stimulus shown belonged to the face- or the scrambled face category (two alternative forced choice). A single trial consisted of an SFM stimulus presentation followed by a long variable inter-stimulus fixation (baseline) interval (on average 9.6 sec) during which responses were given. Prior to the present fMRI experiment, all elderly and MCI patients performed a similar psychophysical task to calibrate the paradigm used here (in preparation).

### ***Localizer runs and task***

During the face localizer scans, participants viewed blocks of stimuli from a given class (faces, places, objects, scrambled images), divided into two runs. Each run consisted of 12 blocks lasting 20 s (20 images, 800 ms duration each, 200 ms inter-stimulus intervals), separated by 10 s fixation baseline intervals. In each block, participants performed a 1-back task to keep stable attention levels. Three repetitions per block were employed. To localize area MT, a single 4.18 min run was acquired in which 30 sec-blocks (6 stimuli, 4.5 sec duration each, 500 ms inter-stimulus-intervals) of coherently moving texture patterns alternated with 30 sec-blocks of the same static pattern. Between blocks there were 10 sec periods of fixation. Participants were instructed to watch passively but attentively.

## **Data analysis**

### ***Behavioural data***

For the statistical analysis of the behavioural data a  $d'$  (dprime) performance measure was computed for face detection in the 9 conditions (3 durations x 3 depth-levels) in each participant [ $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$ ]. Performance data were entered into repeated-measures general linear models (GLMs) with within-subject factors stimulus type, depth, duration and between-subject factor group. Where applicable, p-values were corrected for non-sphericity using the Greenhouse-Geisser or Huynh-Feldt correction.

### ***fMRI data***

#### ***Preprocessing***

The data were pre-processed and analyzed using Brainvoyager QX 2.3 (Brain Innovation, Maastricht, the Netherlands). Slice-scan-time correction, head-motion-correction as well as temporal high-pass filtering were applied and subsequently anatomical and functional data were spatially normalized to the Talairach coordinate system (Talairach and Tournoux 1988). The sinc interpolation option was used to transform each brain into the size of the standard Talairach brain using manually specified reference points. Within-run movement exceeding more than 3 mm led to the exclusion of a run for further analysis. In addition, spatial smoothing (8 mm) of individual datasets was applied only for the group analysis.

### *ROI analyses*

Regions of interest (ROIs) were defined for each participant individually by face and motion localizer scans. The ROIs responding preferentially to faces (FFA, OFA, STS) were defined by the contrast faces > objects, places and scrambled images ( $p < 10^{-4}$ , uncorrected) and were restricted to a maximum cluster size of 125 voxels, to achieve optimal statistical face sensitivity (Fox et al. 2009). Motion sensitive area MT was identified by the contrast motion > no motion ( $p < .005$ , Bonferroni corrected). The statistical thresholds used are based on conventional definitions. Talairach coordinates of the localized ROIs were consistent with previous studies (see Table 2).

To examine fMRI responses in the face-selective ROIs and in MT, z-scored beta weights were extracted for each condition from unsmoothed files of each participant. Subsequently, beta weights were subjected to a separate repeated-measures ANOVA for each ROI. We performed 2 different analyses: First, to gain insight into face perception from SFM in our 3 participant groups, we pooled across depth levels (but not durations) and compared ROI activations for intact and scrambled SFM face conditions. Therefore, we computed repeated-measures ANOVAs for each duration using stimulus type (scrambled vs intact) as within-subject factor and participant group (3 levels) as between-subject factor. Second, to understand the overall effect of the depth manipulation on the fMRI response we pooled data from intact and scrambled SFM face stimuli in the face-selective regions FFA, OFA and STS as well as in area MT and computed separate GLMs per participant group. Within-subject factors were depth level and duration. The magnitude of the depth effect was estimated using the partial eta squared index (partial  $\eta^2$ ), quantifying the percentage of overall variance explained by the depth factor alone after factoring out effects of other variables.

### *Linear discriminant analysis*

Based on the effects revealed in the ROI analysis of right and left FFA/OFA, a linear discriminant function was used to test whether group membership (MCI vs healthy elderly controls) could be predicted from a linear combination of face and depth-sensitivity scores. As face-sensitivity index we used the difference between beta values of faces and scrambled faces in the medium duration condition of right FFA and OFA. The depth-sensitivity index was determined by the slope computed from the beta values associated with the 3 depth levels in left FFA and left OFA.

### *Whole brain analysis*

Random effects (RFX) whole brain analysis was used to investigate between group activation differences. As in the behavioural data analysis, comparisons were restricted to the investigation of effects due to healthy ageing (young vs healthy elderly) and the specific impairments in MCI compared with age-matched controls (healthy elderly vs MCI). Statistical maps were projected onto an 'average brain' that was computed by averaging cortices from all participants. First, to identify regions that show differences between groups during SFM perception (overall task

performance), we computed difference maps for all SFM conditions pooled. Second, to answer the question whether there are group differences in brain regions that are recruited to *successfully* perform the SFM task, we also computed group difference maps for the contrast correct-incorrect trials. Resulting difference maps were first thresholded at  $p < 0.02$ , uncorrected for multiple comparisons and were then submitted to cluster-level statistical threshold estimation based on Monte Carlo simulations with 1000 iterations. The minimum cluster size that survived the thresholding was then used as cluster-size threshold of the source statistical map corresponding to a multiple comparison correction of  $p = 0.05$  at the cluster-level.

Table 2. Talairach coordinates and cluster sizes of ROIs

Region	N	Talairach coordinates			Cluster size
		Mean			
	Total (young,elderly,MCI)	x	y	z	Mean (SD)
rFFA	35 (11,12,12)	39	-45	-21	112 (21.7)
lFFA	35 (11,12,12)	-40	-48	-19	109 (22.8)
rOFA	35 (11,11,11)	39	-70	-13	110 (23.4)
lOFA	26 (8, 9, 9)	-40	-73	-17	111 (17.5)
rSTS	30 (11,9,10)	49	-53	7	91 (36.4)
lSTS	27 (9, 9, 9)	-51	-55	7	92 (28.7)
rMT	32 (12,10,10)	43	-68	1	119 (15.2)
lMT	32 (12,10,10)	-45	-72	-1	123 (8.2)

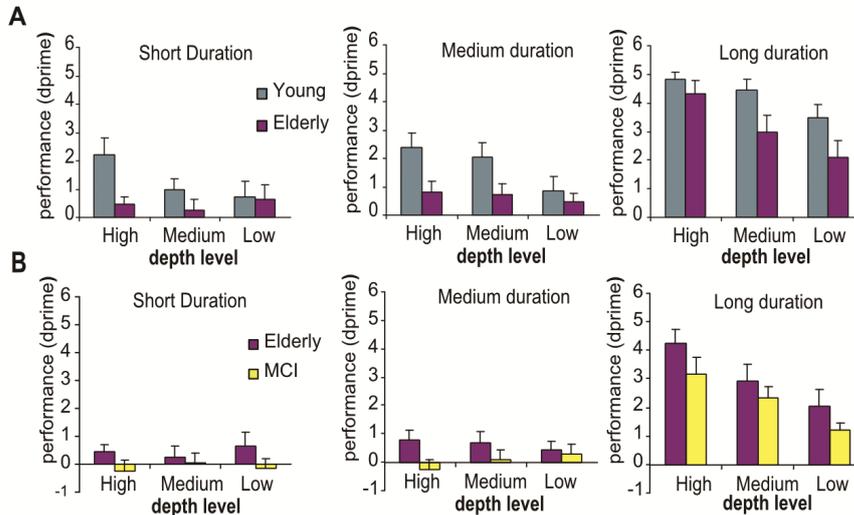
## Results

### Behavioural results

A  $d'$  performance measure was entered into 2 separate repeated measures ANOVAs with within-subject factors depth, duration and between-subject factor group. The first ANOVA compared face detection performance in healthy young and healthy elderly (i.e. effects of healthy ageing, Fig. 2A). This analysis revealed significant main effects of depth ( $F_{(2,46)} = 17.987$ ,  $p = 0.001$ ), duration ( $F_{(2,46)} = 92.971$ ,  $p = 0.001$ ) and group ( $F_{(1,23)} = 4.560$ ,  $p = .044$ ) as well as a significant 3-way interaction between the factors depth, duration and group ( $F_{(4,92)} = 2.766$ ,  $p = .032$ ). Post-hoc tests per duration indicated improved performance at higher depth levels in the medium and long duration conditions ( $F_{(2,46)} = 3.921$ ,  $p = .027$ ;  $F_{(2,46)} = 24.419$ ,  $p = .001$ , respectively), but not in the short duration condition ( $p > 0.5$ ). A significant effect of the between-subject factor group was only found in the *medium* duration condition ( $F_{(1,23)} = 5.025$ ,  $p = .035$ ), indicating better performance in the young compared to the elderly group.

The 2<sup>nd</sup> ANOVA, comparing elderly and MCI participants (i.e., effect of pathological ageing, Fig. 2B), revealed a significant main effect of depth ( $F_{(2,48)} = 4.206$ ,  $p = .021$ ), duration ( $F_{(2,48)} = 63.971$ ,  $p < 0.001$ ) and group ( $F_{(1,24)} = 4.493$ ,  $p = .045$ ) as well as an interaction between depth and duration ( $F_{(4,96)} = 8.980$ ,  $p < .001$ ). Post-hoc analyses per duration showed a significant main effect of depth only for the longest duration ( $F_{(2,48)} = 27.276$ ,  $p < 0.001$ ). Interestingly, also healthy

elderly and MCI participants only showed statistically different performance in the *medium* duration stimuli ( $F_{(1,24)} = 4.481, p = .045$ ). These results suggest that the medium duration condition is best suited for discriminating between participant groups.



**Figure 2. Psychophysics SFM task performance in the young, elderly and MCI.** Bargraphs indicate SFM task performance ( $d'$  performance measure) for the 3 durations and 3 depth levels. **A** Analysis including the young and elderly participants revealed that depth influenced task performance significantly in the medium and long duration conditions but not in the short durations. Performance between groups differed significantly in the medium duration condition for which young participants outperformed the elderly. **B** Investigating differences in SFM task performance between elderly and MCI participants revealed that depth influenced performance only in the longest duration condition while it had no effect in the medium and short duration conditions. Interestingly, differences in performance between healthy elderly and MCI participants were again only significant in the medium duration condition.

### fMRI: ROI Analysis

Face-selective clusters were located in the bilateral middle fusiform gyrus (FFA), bilateral inferior occipital gyri (OFA) and bilateral superior temporal sulci (STS). Motion-selective area MT was localized in bilateral medial temporal sulci (Table 2).

### *Differential activation patterns to SFM faces in ageing and MCI*

Here we intended to study differences in SFM face perception among groups in areas FFA, OFA, STS and MT. A repeated measures GLM with the within-subject factors stimulus type (intact vs scrambled face) and between-subject factor group (young, elderly, and MCI) was performed after pooling across the 3 depth conditions. Since differences in task performance between participant groups were found to be critically dependent on stimulus duration (Fig. 2B), we here computed separate a priori planned ANOVAS for each duration condition. As the most interesting and consistent effects (behaviourally and in the ROI analysis) were

found in the *medium* duration, Figure 3 shows only the results of this condition. For comparison and completeness results for both left and right ROIs are depicted.

### FFA

In right FFA, the short and long duration conditions did not reveal any significant stimulus type or group effects ( $p > .05$ ), but the medium duration condition (Fig. 3A, left column) showed differential patterns of category specificity in the 3 groups (interaction group  $\times$  stimulus type:  $F_{(2,32)} = 6.734, p = .004$ ). Main effects of stimulus type and group were not significant ( $p > .05$ ). Post-hoc tests per group revealed a significant SFM face-sensitive effect in the young group ( $F_{(1,10)} = 5.797, p = .037$ ), shown by significantly larger activation for faces compared to scrambled control faces. The elderly group showed no evidence for SFM face sensitivity in this region ( $F_{(1,11)} = .555, p = .472$ ). Interestingly, in the MCI patients, the data pattern reversed compared to the young group, showing significantly larger activation for scrambled faces than for intact faces ( $F_{(1,11)} = 7.320, p = .02$ ).

Analysis in left FFA (Fig. 3B, left column for data in medium duration condition) revealed neither a significant main effect of stimulus type nor significant interactions between stimulus type and group in any of the durations ( $p > .05$ ). A significant effect of the between-subject factor group was found in the short ( $F_{(2,32)} = 3.528, p = .041$ ) and medium duration conditions ( $F_{(2,32)} = 5.560, p = .008$ ) (Fig. 3B, left column). Post-hoc testing suggested decreasing activation in the healthy elderly and MCI compared to the young group, but after Bonferroni correction only reduced activity in MCI relative to young proved significant ( $p = .004$  and  $p = .042$ , for the short and medium duration conditions respectively).

### OFA

Similar to right FFA, also right OFA revealed an absence of significant effects in the short and the long duration conditions ( $p > .05$ ), while in the medium duration condition (Fig. 3A, right column) face-related activation differed significantly between the groups (interaction stimulus type  $\times$  group:  $F_{(2,30)} = 5.361, p = .011$ ). Main effects of stimulus type and group did not reach significance ( $p > .05$ ). Testing for a main effect of stimulus type separately in each group revealed significantly greater activity for faces than for scrambled controls in the young group ( $F_{(1,10)} = 6.983, p = .025$ ), a lack of face sensitivity in the elderly group ( $F_{(1,10)} = .072, p = .793$ ) and smaller activity for faces than scrambled faces in the MCI group ( $F_{(1,10)} = 6.918, p = .025$ ).

Left OFA showed no significant main effects or interactions in the medium and long durations, although the medium duration condition (Fig. 3B, right column) showed a trend towards a data pattern similar to the one observed in right FFA and right OFA (interaction stimulus type  $\times$  group:  $F_{(2,23)} = 2.600, p = .096$ ). The short duration showed a significant interaction between stimulus type and group ( $F_{(2,23)} = 6.454, p = .006$ ), which was due to activation for faces being smaller than for scrambled faces in the elderly ( $F_{(1,8)} = 6.093, p = .039$ ) while the other groups showed no stimulus specificity.

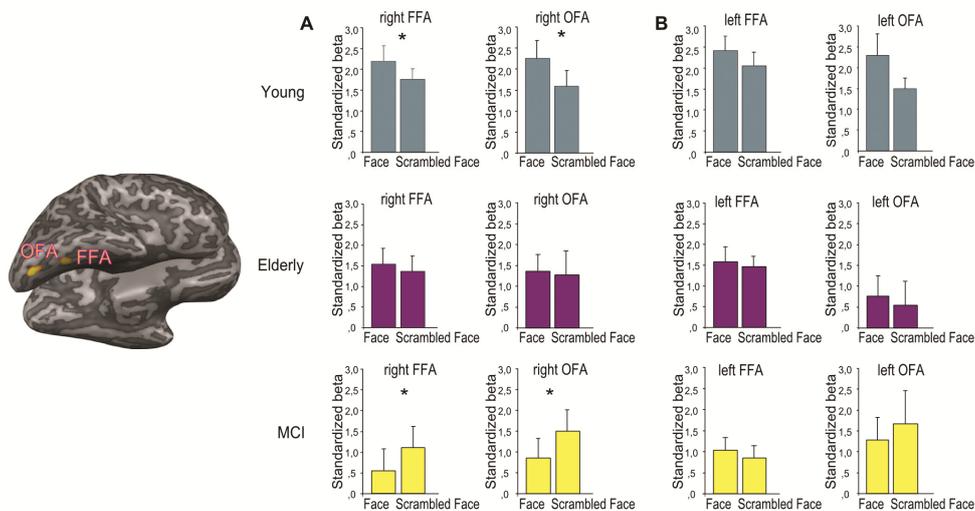
### STS

ROI Analysis in STS revealed no face-sensitive effects, no group effects and no significant interactions ( $p > .05$ ), indicating that this region did not contribute significantly to face-sensitive SFM processing in the present task in any of the groups (but see also whole-brain analysis results).

### MT

We found significant effects of stimulus type (larger activation to scrambled faces compared with faces) in the short duration conditions for left MT ( $F_{(1,29)} = 4.318, p = .047$ ) and right MT ( $F_{(1,29)} = 9.646, p = .004$ ) but not for the other durations ( $p > 0.5$ ). There were no group effects and no significant interactions ( $p > .05$ ).

In sum, healthy ageing was characterized by the disappearance of face-sensitive responses in both right FFA and right OFA. In MCI, right FFA and OFA showed an extreme decline of face-sensitive responses witnessed by *greater* activation for scrambled than for intact faces, suggesting an increased sensitivity to low-to-intermediate level stimulus features at the detriment of sensitivity for the higher-level features that define a face. These fMRI findings are specific to the medium stimulus duration, and therefore likely to be related to differences in task performance among the participant groups at that duration (Fig. 2).



**Figure 3. Face sensitivity in areas FFA and OFA** **A** A ROI analysis in face-sensitive areas FFA and OFA in the right hemisphere showed significant differences in face sensitivity between the 3 participant groups in the *medium* duration condition. While right FFA and right OFA were selective to SFM faces in the young group, in the elderly group activation between faces and scrambled faces did not differ. Interestingly, the MCI patients showed higher sensitivity to scrambled control faces. **B** Analysis of FFA and OFA in the left hemisphere revealed no significant category specific effects in any of the groups, however a trend towards the data pattern found in the right FFA and OFA was also observed in left OFA. Bargraphs indicate standardized beta values of the face and scrambled face conditions (depth levels are pooled) in the medium duration condition. Error bars indicate the SEM. Asterisks indicate significant differences between conditions at the 0.05 level. For illustration, individually localized FFA and OFA are shown on a reconstructed inflated cortical surface of one representative participant.

### **Increased sensitivity to depth from SFM perception in MCI patients**

To test the idea of increased/abnormal susceptibility to stimulus features in MCI as an explanation for the reduced category specificity in the previous analysis, we compared sensitivity to stimulus depth among the 3 groups. For this analysis, face and scrambled face conditions were pooled and depth (high, medium, low) and duration (short, medium, long) were included as within-subject factors in the analysis. Separate ANOVAs that had been planned a priori were computed for each group (young, elderly, MCI). Since we aimed to test depth-sensitivity, we additionally computed partial eta squared values ( $\eta^2$ ), which quantify the percentage of overall variance explained by the depth factor alone (effect size) after factoring out effects of duration (see Materials and Methods) (Fig. 4).

#### *FFA*

In the right FFA, none of the groups showed a significant main effect of depth ( $p > .05$ ; partial  $\eta^2$  young = .054, partial  $\eta^2$  elderly = .060, partial  $\eta^2$  MCI = .10; effect sizes  $\eta^2$  shown in Figure 4A). There were significant main effects of duration in all groups ( $p < .05$ ), but no significant interactions between depth and duration in any of the groups ( $p > .05$ ). On the other hand, analysis in left FFA revealed a surprisingly strong sensitivity to stimulus depth, which was present only in the MCI group ( $F_{(2,22)} = 10.182$ ,  $p = .001$ , partial  $\eta^2 = .481$ ) but not in the young and elderly control groups ( $p > .05$ , partial  $\eta^2$  young = .190, partial  $\eta^2$  elderly = .066) (Fig. 4B). Duration had a significant effect in all 3 groups ( $p < .05$ ) while none of the interactions reached significance ( $p > 0.5$ ).

#### *OFA*

In the right OFA (Fig. 4C), only the young group showed a significant depth effect ( $F_{(2,20)} = 5.689$ ,  $p = .011$ ) partial  $\eta^2 = .362$  in the absence of any depth effects in the elderly and MCI ( $p > .05$ , partial  $\eta^2 = .085$  and partial  $\eta^2 = .083$ , respectively). Duration was significant in all 3 groups ( $p < .05$ ) but did not interact with depth in any of the groups ( $p < .05$ ). In the left OFA (Fig. 4D), similar to left FFA (Fig. 4B), there was neither a significant depth effect ( $p > .05$ , partial  $\eta^2 = .290$ ) in the young group, nor in the elderly group ( $p > .05$ , partial  $\eta^2 = .114$ ), but there was a strong depth effect in the MCI group ( $F_{(2,16)} = 6.859$ ,  $p = .007$ , partial  $\eta^2 = .462$ ). In all groups the factor duration had a significant main effect ( $P < .05$ ) but did not interact significantly with depth ( $p > .05$ ).

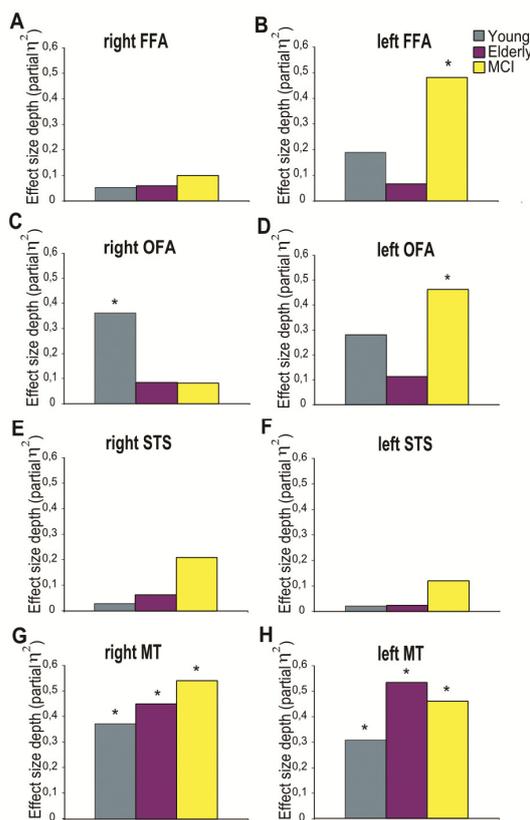
#### *STS*

In right and left STS, there were no significant main effects of depth ( $p > .05$ ) (Fig. 4E & 4F) while the factor duration was significant in all groups ( $P < .05$ ). However, in the right STS of the MCI, the factors depth and duration interacted ( $F_{(4,36)} = 2.910$ ,  $p = .035$ ) indicating a sensitivity to depth only in the short ( $F_{(2,18)} = 6.279$ ,  $p = .009$  partial  $\eta^2 = .411$ ) and long duration conditions ( $F_{(2,18)} = 4.209$ ,  $p = .032$  partial  $\eta^2 = .319$ ) but not in the medium duration ( $p > .05$ , partial  $\eta^2 = .033$ ). This indicates abnormalities in depth processing of SFM stimuli also in STS of the MCI group.

**MT**

Area MT in all 3 groups revealed a strong sensitivity to stimulus depth in the right hemisphere (young:  $F_{(2,22)} = 6.669, p = .005, \text{partial } \eta^2 = .377$ ; elderly:  $F_{(2,18)} = 7.522, p = .004, \text{partial } \eta^2 = .455$ ; MCI:  $F_{(2,18)} = 10.846, p = .001, \text{partial } \eta^2 = .547$ ) (Fig. 4G) as well as in the left hemisphere (young:  $F_{(2,22)} = 4.895, p = .017, \text{partial } \eta^2 = .308$ ; elderly:  $F_{(2,18)} = 10.283, p = .001, \text{partial } \eta^2 = .533$ ; MCI:  $F_{(2,18)} = 7.684, p = .004, \text{partial } \eta^2 = .461$ ) (Fig. 4H). Stimulus duration was significant in all 3 groups ( $p < .05$ ) and did not interact with depth. These results suggest intact mechanisms for extracting depth from motion in area MT in all 3 participant groups.

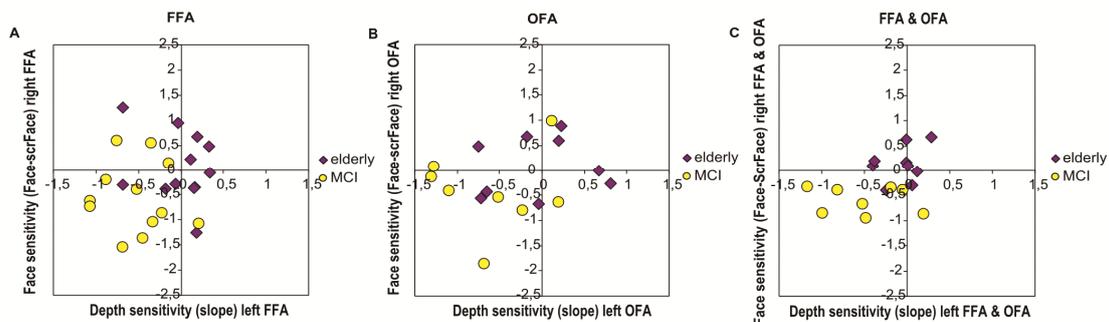
In sum, we found a high sensitivity to depth in the left FFA and left OFA of MCI patients that was absent in the young and healthy elderly participants. These results suggest that the decreased sensitivity for faces in right FFA and OFA might be in part related to increased sensitivity in the left hemisphere to lower-level aspects of SFM stimuli.



**Figure 4. Depth sensitivity in areas FFA, OFA, STS and MT** **A&B** Region-of-Interest analysis in FFA revealed an unusually large sensitivity to variations in stimulus depth (large partial eta squared) in the left but not the right FFA of MCI patients. The young and the elderly control groups showed no significant depth effects in this region. **C&D** A similar lateralization pattern was also observed in area OFA. ROI analysis in the left OFA showed high depth-sensitivity in the MCI but not in the young and elderly. Instead, the young were sensitive to depth in the right OFA. **E&F** In both left and right STS no overall depth-sensitivity was found, however, bargraphs indicate a trend towards higher depth-sensitivity in the MCI patients compared to the control groups. **G&H** All 3 groups exhibited a large depth-sensitivity in bilateral MT. Bar graphs indicate the effect size (partial eta squared) of the depth variable for all 3 groups pooled over duration and stimulus type. Asterisks indicate significant main effects of depth at the 0.05 level.

### **Classifying MCI and elderly control groups based on depth- and face-sensitivity**

To assess the diagnostic utility of the effects revealed in the above reported ROI analysis, we tested whether group membership (MCI vs healthy elderly controls) could be predicted from face and depth-sensitivity scores (based on beta values, see Methods) from FFA and OFA using a linear discriminant function (for a similar application of linear discriminant analysis (LDA) see Ihssen et al. (2011)). Thus, each participant was scored on both the face and depth index leading to scatter plots for the healthy elderly and the MCI characterized along two dimensions. Figure 5 depicts results for FFA (Fig. 5A), OFA (Fig. 5B) and combined data from FFA/OFA (Fig. 5C). The discriminant function is based on a linear combination of the two predictor variables that best separates the groups. Results for FFA showed larger standardized canonical discriminant function coefficients for the depth (.886) than for the face index (.628) indicating that the depth index had greater discriminating ability. The function classified MCI and elderly controls with 83.3 % accuracy (*Wilks'  $\lambda$*  = .543,  $\chi^2$  = 12.816, *df* = 2, *p* = .002). In OFA, the face index had slightly higher discriminating ability than the depth index (standardized canonical discriminant function coefficients .731 and .693, respectively) while the model predicted group membership with 70.6% accuracy (*Wilks'  $\lambda$*  = .647,  $\chi^2$  = 6.104, *df* = 2, *p* = .047). A linear discriminant function including all 4 predictors, i.e. depth and face indices from both FFA and OFA yielded the highest accuracy (*Wilks'  $\lambda$*  = .343,  $\chi^2$  = 13.910, *df* = 4, *p* = .008, 88.2 % of cases correctly classified). Taken together these findings indicate that in the present study, MCI patients and healthy elderly participants could be classified with high accuracy based on face and depth-sensitivity indices in face sensitive regions FFA and OFA.



**Figure 5. Face- and depth-sensitivity indices discriminate between MCI patient and elderly control group** A linear discriminant analysis indicated that face and depth sensitivity patterns in FFA and OFA were highly predictive of group membership. Scatterplots indicate face indices in right FFA/OFA plotted against depth indices in left FFA/OFA for each participant in the healthy control and MCI group. **A** Using face and depth indices in FFA, the distributions for the healthy elderly and the MCI could be separated correctly in 83.3 % of the cases **B** In OFA the model classified between healthy elderly and MCI with 70.6% accuracy. **C** A combination of depth and face indices from all 4 ROIs (right FFA/OFA and left FFA/OFA) as predictors yielded a prediction accuracy of 88.2 %.

## **fMRI: Whole brain analysis**

### ***Between-group comparisons: Overall-activation differences***

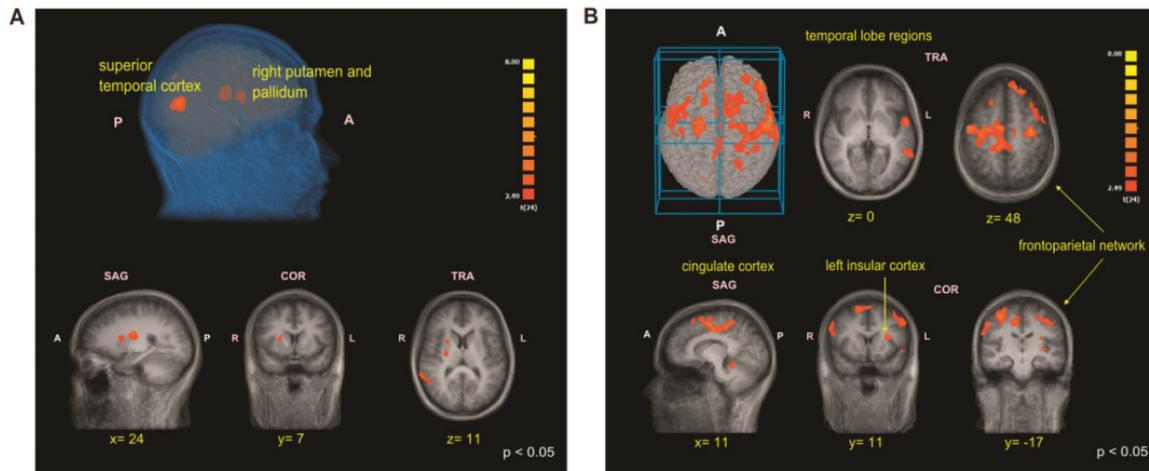
Comparing overall activation between the young and the healthy elderly group during the SFM face categorization task pooled over all conditions revealed more activity in the elderly compared to the young in a bilateral superior parietal network, in bilateral inferior temporal and parahippocampal gyri, and in predominantly right-lateralized frontal regions. Conversely, the elderly showed significantly less activation than the young in a small patch in the posterior cingulate cortex ( $p < .05$ , corrected).

The overall activation comparison between healthy elderly and MCI group revealed stronger activation in the MCI than in healthy elderly in the right putamen and right pallidum as well as in an area in the right superior temporal lobe ( $p < .05$ , corrected) (Fig. 6A). There were no areas that showed less activation in the MCI than in the elderly.

### ***Between-group comparisons: Regions involved in correct task performance***

Differences in the network involved in correct task performance between young and elderly control participants were found in the right superior temporal pole, which was more active in the elderly compared to the young participants. On the other hand, elderly showed less activation than the young in bilateral inferior parietal regions as well as in confined bilateral inferior occipital regions presumably corresponding to OFA ( $p < .05$ , corrected).

Activation associated with correct task performance (correct > incorrect trials) also differed significantly between elderly control participants and the MCI patients. The MCI group displayed more activation in bilateral medial parietal/cingulate cortex, the left anterior and medial temporal cortex, the left parahippocampal cortex and a predominantly right-lateralized fronto-parietal network including a region in the left insular cortex and inferior parietal cortex, relative to the elderly control group ( $p < .05$ , corrected) (Fig. 6B). There were no areas that were less activated in the MCI compared to the elderly for this contrast.



**Figure 6. Whole brain RFX analysis: Differences between elderly controls and MCI** **A** Group difference maps between elderly controls and MCI for overall activation differences during SFM perception showed areas in the right superior temporal lobe as well as parts of the right putamen and the right pallidum that were significantly more active in the MCI. Patches with lower colour intensity indicate areas located inside the brain viewed through a low opacity illustration of the cortex. Right hemisphere activity is also shown in sagittal (SAG), coronal (COR), and transversal (TRA) slices. **B** Contrasting correct and incorrect responses and comparing the resulting activated areas between elderly control and MCI group revealed that in correctly solving the SFM task, the MCI recruited additional areas including a bilateral slightly more right-lateralized frontoparietal network, the medial parietal/cingulate cortex, a parahippocampal area and regions in the anterior and medial temporal lobe ( $p < .05$ , corrected). Differential activity is shown in a brain surface view, with additional views on sagittal (SAG), coronal (COR), and transversal (TRA) slices. Slice locations are given in Talairach coordinates.

## Discussion

In the present study, we aimed to detect early alterations in visual brain function in amnesic MCI, a condition associated with likely conversion to AD. To that goal, we used SFM stimuli enabling us to study both dorsal and ventral stream visual functions using concurrent behavioural and fMRI measurements.

Our findings indicate a decline in the ability to extract face information from short lived motion cues in healthy ageing and particularly in MCI, witnessed by behavioural measures and fMRI activation in temporal lobe areas. Young participants showed selective fMRI activation to SFM faces compared to scrambled control stimuli in right FFA/OFA and low sensitivity to stimulus depth modulations in the left FFA/OFA. In healthy elderly participants this face-sensitive processing in the right FFA/OFA was greatly reduced while an essentially normal low sensitivity to local features in the left hemisphere was retained. Intriguingly, in the MCI we found higher fMRI responses to scrambled than to intact faces in right FFA/OFA associated with a large sensitivity for low-level stimulus features (depth) in left FFA/OFA. This reversed sensitivity pattern in FFA and OFA marks a fundamental distinction between healthy ageing and MCI participants that is closely associated with their task performance.

Additional findings included overall stronger activation in MCI than in healthy elderly in the right putamen, right pallidum and an area in the right superior temporal lobe, as well as increased activation in a broad network of areas during correct as opposed to incorrect performance. Our findings on healthy and pathological ageing in MCI will be first placed in the context of hemispheric asymmetries in normal face perception, after which contributions of impairments in dorsal-ventral integration and attention to MCI will be discussed.

### **Hemispheric asymmetries in face perception in normal and MCI participants**

FFA, OFA and STS constitute the core regions of the face processing system in healthy adults (Kanwisher and Yovel 2006). Area FFA has also been shown to be selectively activated in response to SFM faces (Kriegeskorte et al. 2003). Here we confirm this finding for right FFA and additionally report SFM face sensitivity in right OFA. However no face-sensitive effects were found in equivalent regions in the left hemisphere. This lateralization of face-sensitive effects is in line with behavioural data indicating stronger face-sensitive effects in the right hemisphere and generally larger right-lateralized face sensitivity in FFA (for a review see (Cabeza and Nyberg 2000)). These findings relate to a more general lateral asymmetry that has been observed in visual processing, having inspired notions of analytic versus configural processing (Bradshaw and Nettleton 1981) or processing high versus low spatial frequencies (Flevaris and Robertson 2011; Ivry and Robertson 1998) in left and right hemispheres respectively. A more recent model proposes that the left hemisphere focuses on isolating features as well as associative relationships between features while the right hemisphere is sensitive to holistic stimulus aspects and conjoining features (Dien 2009). In the context of the present study this would imply that the left FFA/OFA extracts face-related features from spatial and depth cues in the SFM stimuli while the right FFA/OFA is involved in assembling visual features for more holistic processing which in turn would lead to a greater face-sensitive fMRI response.

While the loss of SFM face sensitivity in right FFA/OFA of the elderly control participants is likely due to 'dedifferentiation', a phenomenon referring to less distinct neural activation in the elderly (Carp et al. 2011; Lee et al. 2011), the fMRI activation patterns in the MCI suggest a fundamental change in the different hemispheric mechanisms for face perception. The stronger activation to scrambled faces compared to faces in the medium duration condition (most performance discriminative) indicates that holistic processing of faces in right FFA/OFA is impaired, presumably contributing to the observed decline in psychophysical task performance. The mechanism underlying this impairment of integrative function may be an over-representation of lower-level stimulus features, which is suggested both by the preference for scrambled compared to intact faces in right FFA/OFA and additionally by the increased sensitivity to stimulus depth in left FFA/OFA. Thus, the absence of the "normal" selectivity pattern in MCI suggests that instead of a holistic approach other, less efficient perceptual strategies were used. While

the alterations in left and right FFA/OFA might reflect independent parallel processes, one might also speculate that they are functionally dependent. In terms of the different roles that have been assigned to left and right hemispheres, the over-emphasized low-level stimulus aspects in left FFA/OFA of the MCI may hinder face-related feature selection and therefore transmit ambiguous output to the equivalent regions in the right hemisphere. This in turn may prevent the conjunction of features and holistic processing in the right hemisphere, leading to higher activity in right FFA/OFA to scrambled rather than to intact face stimuli. In this model, left FFA/OFA would operate as an interface, receiving input from other areas (e.g. STS) and providing output to right FFA/OFA. The strong functional connectivity between corresponding face processing regions across hemispheres (Davies-Thompson and Andrews 2011) is in line with this view. However, these speculative ideas about possible interactions between left and right FFA/OFA/STS remain to be tested in future studies using functional connectivity analysis, as the present study was not set up to do so.

Results of the present SFM fMRI paradigm mark a fundamental distinction between MCI and healthy elderly participants. Strikingly, a linear discriminant function using face-sensitivity indices in right FFA/OFA and depth-sensitivity indices in left FFA/OFA permitted a classification of healthy elderly and MCI with a high degree of precision (88.2%). While these results are promising, it should however be acknowledged that their interpretation is constrained by the limited number of participants (N=38 participants: 12 young, 13 elderly, 13 MCI). Therefore, future studies with larger sample sizes are required to establish the consistency of the effects before their diagnostic value can be assessed and clinical use can be contemplated.

### **Impaired dorsal-ventral cross-talk during face perception in MCI**

Psychophysical studies have indicated specific deficits in SFM perception in AD (Kim and Park 2010; Rizzo 2000) and more recently also in MCI (Lemos et al. 2012). Despite reports of impairments at motion perception tasks in AD (Kavcic et al. 2011) as well as alterations in fMRI activation patterns related to motion processing (Thiyagesh et al. 2009), the present study showed normal patterns of fMRI activation in MT of the two elderly groups. Hence, the observed SFM impairments in our study are unlikely to primarily reflect motion processing deficits. Instead, since the integration of information from dorsal and ventral visual streams is crucial in SFM perception, we suggest that the mechanisms of dorsal-ventral integration might be impaired in MCI. The superior temporal cortex has long been hypothesized to play a major role in dorsal-ventral interaction. From monkey anatomical studies it is known that the STS/STG receives inputs from both streams (Morel and Bullier 1990; Seltzer and Pandya 1978) and has direct anatomical connections to the putamen, caudate nucleus and the pulvinar (Yeterian and Pandya 1998). Significantly, we found an overall hyper-activation during SFM perception in the MCI in the superior temporal cortex as well as the right putamen and right pallidum. Hyperactivation in the putamen of the MCI may result from microstructural damage in the right caudate (Cherubini et al. 2010), a structure that together with the

putamen forms the dorsal striatum. These findings hence suggest overactivations in MCI that may underlie suboptimal integrative processing in dorsal/ventral conjunction regions, and may be related to the impaired temporal lobe face perception mechanisms in this group. It is conceivable that deficient integration causes an overemphasis of low-level stimulus aspects to the detriment of more global integrative processes in the face processing network.

### **Contributions of attention to SFM face perception in MCI**

Face processing mechanisms are heavily influenced by attention processes. Object specificity in several visual areas including the FFA (Murray and Wojciulik 2004; Reddy and Kanwisher 2007) has been shown to increase when the object category is attended. In AD, selective attention is known to be affected early in the course of the disease (Baddeley et al. 2001). Therefore, in the present study an attentional bias towards predominantly local stimulus features at the detriment of configural processing in right FFA/OFA might underlie the observed pattern of results in MCI. However, results from our whole brain analysis provide strong support for an alternative interpretation, suggesting that attention may have been allocated more strongly in MCI to compensate for an automatic emphasis on lower-level features in SFM face processing in the temporal lobe. This interpretation is confirmed by increased activation in MCI during correct task performance in the medial parietal/cingulate cortex, a fronto-parietal network including the insula, and inferior parietal regions. Several frontal and parietal regions have been shown to be involved in attention (Corbetta and Shulman 2002) and the insula and the medial cingulate cortex are known to play a role in environmental monitoring and response selection (Taylor et al. 2009). In a highly relevant study, Grady et al. (2003) proposed that AD patients compared to controls recruit additional unique neural networks that may indicate compensatory mechanisms for gray matter losses. Hence, also in the present study the hyperactivations in regions involved in attention and executive control may be a mechanism to maintain adequate performance levels in the setting of early brain pathology.

Our study design, focusing on performance-related changes instead of resting-state alterations, adds to previous fMRI studies by using a relevant cognitive task (perception of motion-defined face stimuli) probing both temporal and parietal mechanisms in healthy ageing and MCI. Our principal finding is an overemphasis of low-level at the cost of global processing during SFM face perception in MCI, which was evident in aberrant patterns of activation in FFA/OFA. Furthermore, MCI was characterized by alterations in dorsal-ventral integrative regions and attention-related networks. Although previous studies have emphasized visual dorsal stream dysfunction in MCI, we found activity patterns of visual ventral stream areas FFA/OFA to be a highly accurate biomarker for MCI. With early diagnosis, pharmacological or cognitive therapies may be one way to halt or slow down progression of the disease (e.g. Bokde et al. 2009). Since in mild stages of the disease a high degree of cortical plasticity might still be preserved, timely application of perceptual and other skill learning paradigms (e.g. Kim and Park 2010) as well as decoded

fMRI neurofeedback (Shibata et al. 2011) may be other particularly promising approaches for reversing/sparing visual and other functions in MCI and preclinical AD.

## References

- Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* 7:270–279.
- Andersen RA, Bradley DC. 1998. Perception of three-dimensional structure from motion. *Trends in Cogn Sci.* 2:222–228.
- Baddeley AD, Baddeley HA, Bucks RS, Wilcock GK. 2001. Attentional control in Alzheimer's disease. *Brain.* 1492–1508.
- Baron J-C, Chételat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F. 2001. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage.* 14:298–309.
- Bokde ALW, Lopez-Bayo P, Meindl T, Pechler S, Born C, Faltraco F, Teipel SJ, Möller H-J, Hampel H. 2006. Functional connectivity of the fusiform gyrus during a face-matching task in subjects with mild cognitive impairment. *Brain.* 129:1113–1124.
- Bokde ALW, Lopez-Bayo P, Born C, Dong W, Meindl T, Leinsinger G, Teipel SJ, Faltraco F, Reiser M, Möller H-J, Hampel H. 2008. Functional abnormalities of the visual processing system in subjects with mild cognitive impairment: an fMRI study. *Psychiatry Res.* 163:248–259.
- Bokde ALW, Karmann M, Teipel SJ, Born C, Lieb M, Reiser MF, Möller H-J, Hampel H. 2009. Decreased activation along the dorsal visual pathway after a 3-month treatment with galantamine in mild Alzheimer disease: a functional magnetic resonance imaging study. *J Clin Psychopharm.* 29:147–156.
- Bozzali M, Filippi M, Magnani G, Cercignani M, Franceschi M, Schiatti E, Castiglioni S, Mossini R, Falautano M, Scotti G, Comi G, Falini A. 2006. The contribution of voxel-based morphometry in staging patients with mild cognitive impairment. *Neurology.* 67:453–460.
- Bradshaw NC, Nettleton JL. 1981. The nature of hemispheric specialization in man. *Behav Brain Sci.* 4:51–63.
- Cabeza R, Nyberg L. 2000. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci.* 12:1–47.
- Carp J, Park J, Polk TA, Park DC. 2011. Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. *Neuroimage.* 56:736–743.
- Cherubini A, Péran P, Spoletini I, Di Paola M, Di Iulio F, Hagberg GE, Sancesario G, Gianni W, Bossù P, Caltagirone C, Sabatini U, Spalletta G. 2010. Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients. *J Alzheimer's dis.* 19:1273–1282.

- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat rev Neurosci.* 3:201–215.
- Cronin-Golomb A, Corkin S, Growdon JH. 1995. Visual Dysfunction predicts cognitive deficits in Alzheimer's disease. *Optom Vis Sci.* 72:168–176.
- Davies-Thompson J, Andrews TJ. 2011. The localization and functional connectivity of face-selective regions in the human brain. *J Vis.* Advance Access published September 23, 2011, doi:10.1167/11.11.647
- Dien J. 2009. A tale of two recognition systems: implications of the fusiform face area and the visual word form area for lateralized object recognition models. *Neuropsychologia.* 47:1–16.
- Dupuis-Roy N, Fortin I, Fiset D, Gosselin F. 2009. Uncovering gender discrimination cues in a realistic setting. *J Vis.* 10:1-8.
- Farivar R, Blanke O, Chaudhuri A. 2009. Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *J Neurosci.* 29:5336–5342.
- Fleisher AS, Sherzai A, Taylor C, Langbaum JBS, Chen K, Buxton RB. 2009. Resting-state BOLD networks versus task-associated functional MRI for distinguishing Alzheimer's disease risk groups. *Neuroimage.* 47:1678–1690.
- Flevaris AV, Robertson LC. 2011. Attentional selection of relative SF mediates global versus local processing: Evidence from EEG. *J Vis.* 11:1–12.
- Folstein MF, Folstein SE, McHugh PR. 1975. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 12:189–198.
- Fox CJ, Iaria G, Barton JJS. 2009. Defining the face processing network: optimization of the functional localizer in fMRI. *Hum brain mapp.* 30:1637–1651.
- Gold BT, Johnson NF, Powell DK, Smith CD. 2011. White matter integrity and vulnerability to Alzheimer's disease: Preliminary findings and future directions. *Acta Biochim Biophys Sin.*
- Grady CL, Maisog JM, Horwitz B, Ungerleider LG, Mentis M, Salerno J, Pietrini P, Wagner E, Haxby J. 1994. Age-related changes in cortical blood flow activation during visual processing of faces and location. *J Neurosci.* 14:1450.
- Grady CL, McIntosh AR, Beig S, Keightley ML, Burian H, Black SE. 2003. Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. *J Neurosci.* 23:986–993.
- Guerreiro M, Fonseca S, Barreto J, Garcia S. 2008. Escala de Avaliacao da Doenca de Alzheimer- EADA. In: Mendonca A, Guerreiro M, editors. *Escalas e Testes na Demência.* 2nd ed. p. 42–58.
- Guerreiro M, Silva A, Botelho M, Leitão O, Castro Caldas A, Garcia C. 2008. Avaliação Breve do Estado Mental (Mini Mental State Examination). In: Mendonca A, Guerreiro M, editors. *Escalas e Testes na Demência.* 2nd ed. p. 33–39.

- Hampel H, Bürger K, Teipel SJ, Bokde ALW, Zetterberg H, Blennow K. 2008. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement.* 4:38–48.
- Ivry R, Robertson L. 1998. *The two sides of perception.* Cambridge, MA: The MIT press.
- Ihssen N, Cox, WM, Wigget A, Fadardi JS, Linden DEJ. 2011. Differentiating Heavy from Light Drinkers by Neural Responses to Visual Alcohol Cue and Other Motivational Stimuli. *Cereb Cortex.* 21:1408–1415.
- Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci.* 361:2109–2128.
- Kavcic V, Vaughn W, Duffy CJ. 2011. Distinct visual motion processing impairments in aging and Alzheimer's disease. *Vision Res.* 51:386–395.
- Kim N, Park J-H. 2010. Learning to perceive structure from motion and neural plasticity in patients with Alzheimer's disease. *Neuropsychologia.* 48:1464–1471.
- Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. *J Neurosci.* 23:1451–1463.
- Lee Y, Grady CL, Habak C, Wilson HR, Moscovitch M. 2011. Face Processing Changes in Normal Aging Revealed by fMRI Adaptation. *J Cogn Neurosci.* 23:3433–3447.
- Lemos R, Figueiredo P, Santana I, Simões MR, Castelo-Branco M. 2012. Temporal Integration of 3D Coherent Motion Cues Defining Visual Objects of Unknown Orientation is Impaired in Amnesic Mild Cognitive Impairment and Alzheimer's Disease. *J Alzheimers dis.* 27:1–12.
- Li Y, Wang Y, Wu G, Shi F, Zhou L, Lin W, Shen D. 2011. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol aging.*
- Liu CH, Collin C a, Farivar R, Chaudhuri A. 2005. Recognizing faces defined by texture gradients. *Percept Psychophys.* 67:158–167.
- Mapstone M, Logan D, Duffy CJ. 2006. Cue integration for the perception and control of self-movement in ageing and Alzheimer's disease. *Brain.* 129:2931–2944.
- Mapstone M, Steffenella TM, Duffy CJ. 2003. A visuospatial variant of mild cognitive impairment: getting lost between aging and AD. *Neurology.* 60:802–808.
- Mendez M, Cherrier M, Meadows R. 1996. Depth perception in Alzheimer's disease. *Percept Mot Skills.* 83:987–995.
- Morel A, Bullier J. 1990. Anatomical segregation of two cortical visual pathways in the macaque monkey. *Vis Neurosci.* 4:555–578.
- Mosimann UP, Mather G, Wesnes KA, O'Brien JT, Burn DJ, McKeith IG. 2004. Visual perception in Parkinson disease dementia and dementia with Lewy bodies. *Neurology.* 63:2091–2096.

- Murray SO, Wojciulik E. 2004. Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci.* 7:70–74.
- Pennanen C, Testa C, Laakso MP, Hallikainen M, Helkala E-L, Hänninen T, Kivipelto M, Könönen M, Nissinen A, Tervo S, Vanhanen M, Vanninen R, Frisoni GB, Soininen H. 2005. A voxel based morphometry study on mild cognitive impairment. *J Neurol, Neurosurg, Psychiatry.* 76:11–14.
- Petersen R, Smith G, Waring S, Ivnik R, Tangalos E, Kokmen E. 1999. Mild cognitive impairment. *Arch Neurol.* 56:303–309.
- Petersen RC. 2004. Mild cognitive impairment as a diagnostic entity. *J Intern Med.* 256:183–194.
- Petersen RC, Doody R, Kurz A, Mohs R, Morris J, Rabins P, Ritchie K, Rossor M, Thal L, Winblad B. 2001. Current concepts in mild cognitive impairment. *Arch neurol.* 58:1985.
- Reddy L, Kanwisher N. 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol.* 17:2067–2072.
- Rizzo M. 2000. Vision and cognition in Alzheimer’s disease. *Neuropsychologia.* 38:1157–1169.
- Sauer J, Ffytche DH, Ballard C, Brown RG, Howard R. 2006. Differences between Alzheimer’s disease and dementia with Lewy bodies: an fMRI study of task-related brain activity. *Brain.* 129:1780–1788.
- Seltzer B, Pandya DN. 1978. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* 149:1–24.
- Shibata K, Watanabe T, Sasaki Y, Kawato M. 2011. Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. *Science.* 334:1413–1415.
- Singh V, Chertkow H, Lerch JP, Evans AC, Dorr AE, Kabani NJ. 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer’s disease. *Brain.* 129:2885–2893.
- Talairach G, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. New York: Thieme.
- Taylor KS, Seminowicz DA, Davis KD. 2009. Two systems of resting state connectivity between the insula and cingulate cortex. *Hum Brain Mapp.* 30:2731–2745.
- Thiyagesh SN, Farrow TFD, Parks RW, Accosta-Mesa H, Young C, Wilkinson ID, Hunter MD, Woodruff PWR. 2009. The neural basis of visuospatial perception in Alzheimer’s disease and healthy elderly comparison subjects: an fMRI study. *Psychiatry Res.* 172:109–116.
- Ungerleider LG, Mishkin M. 1982. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. *Analysis of visual behavior.* Cambridge (MA): MIT press. p. 549–586.
- Wechsler D. 1987. WMS-R Wechsler memory scale - Revised manual. San Antonio, TX: Psychological Corporation.

Yamasaki T, Muranaka H, Kaseda Y, Mimori Y, Tobimatsu S. 2012. Understanding the Pathophysiology of Alzheimer's Disease and Mild Cognitive Impairment: A Mini Review on fMRI and ERP Studies. *Neurol Res Int.* 2012:719056.

Yeterian EH, Pandya DN. 1998. of the Superior Temporal Region in Rhesus Monkeys. *J Comp Neurol.* 402:384–402.



## EPILOGUE

Seeing the world in 3D is part of our everyday lives. Without 3D, it would be impossible to perform figure ground segregation, to recognize and grasp objects, and to move around in our environments. Moreover, a reduction of the repertoire of 3D cues to our disposal by just one (e.g. disparity) is already sufficient to cause perceptual problems, and reduces the impression of things being ‘real’. This surely explains the growing popularity of technologies that add stereo-depth to images seen in cinemas, on television, or in video games. The development of these features constitutes an immense technical challenge and are therefore often associated with high costs (Kalva et al. 2006; Kawai, 2002). Remarkably, given the significant amount of computations involved in creating 3D images, our visual system seems to perform this task seemingly without effort (and without cost!) on a daily basis.

The present thesis focused on the recovery of 3D structure from shading and motion visual image cues as well as their underlying neural mechanisms in human observers. The resulting findings, their implications for current models of 3D visual perception as well as for the empirical literature will briefly be highlight here. Moreover, some directions for future research will be suggested.

### ***Topic 1: Processing of multiple visual image cues for the perception of simple 3D forms***

In our psychophysical study, we employed a task in which participants judged the curvature of simple objects, defined by shading, motion, or combinations of both cues. We found that 3D shading and motion cues did not combine in a statistically optimal fashion according to Maximum Likelihood Estimation (MLE). Instead, the cues showed a form of suppressive competition, as shown by reduced reliability of curvature judgments in combined compared to single cue conditions. Moreover, the large dominance of the shading over the motion cue in the consistent combined condition could not be explained by differences in single cue reliabilities, suggesting that MLE is not applicable to any arbitrary choice of cue pairs.

In line with ideas of Bayesian models of cue integration (e.g. Knill 2007), we suggest that several ecological factors may be important in determining whether some cues combine in a cooperative fashion while others do not. Prior knowledge about the world, where fine curvature of static objects is rarely described by motion cues, might have resulted in an experience-based suppression of the motion cue in the combined percept. Moreover, stimulus ambiguity may be an important factor that puts constraints on cue integration (Bulthoff and Mallot 1987). Both shading and motion cues studied here, are known to be ambiguous with respect the coarse judgments of convexity/concavity required for the present task (Orban 2011; Lovell and Bloj 2012). However, coarse determination of an object’s 3D structure based on single cues may be a precondition for cue combination to become cooperative. Otherwise, fine 3D curvature judgments may be

suppressed by the visual system. Note that in our experiments we randomized the position of the light source in the shading cues to minimize the contributions of local aspects of the image's greyness distribution to performance. It is possible however, that this manipulation hampered the disambiguation of the coarse sign of the curvature whereas a fixed light source shining from above could have helped to disambiguate the shading cue. Therefore, it might be interesting to compromise on the control of local luminance gradients, and combine motion cues with a shading cue in which the position of the light source is kept constant. In such an experiment, cue integration may in fact become cooperative and follow MLE, instead of resulting in the suppressive competition mechanism we observed here.

In chapter 2, we aimed to extend the psychophysical findings of the previous chapter by using fMRI in combination with a similar curvature discrimination task. We found further psychophysical support for a lack of cooperative integration and instead found a weighted averaging of shading and motion cues in combined conditions. The fMRI data, however, were more difficult to interpret. We found no parametric relationship between stimulus curvature (defined by either single or combined cues) and fMRI signal in the group average, which was due to individual participants showing idiosyncratic relationships between fMRI signal and curvature. Interestingly, fMRI-based curvature sensitivity in combined cue conditions could be predicted from a combination of fMRI-based curvature sensitivities in single cue conditions in a right IPS region. However, neither the single cue nor the combined cue fMRI responses correlated with psychophysical data collected in respective conditions. Although we found some correlation between fMRI-based single cue curvature sensitivities in right IPS and behavioural performance as well as RTs in combined cue conditions, these findings are hard to interpret at this stage. The work described in this chapter is work in progress. In the future, multi-voxel pattern analysis (MVPA), which instead of spatial averaging of neuronal population responses uses large-scale multivariate patterns of BOLD signals, may be able to shed more light on the data we have collected. Effectively, MVPA has shown parametric dependencies of the fMRI signal on disparity-defined depth (Preston et al. 2008) and may be able to elucidate the relationship between performance and fMRI activation patterns in the present dataset. Specifically, it would be interesting to not only test the decoding of curvature related information in single cue conditions, but to also test whether activity patterns elicited by combined cue stimuli would be a weighted average of activity patterns elicited in single cue stimuli. This technique has been successfully applied to test biased competition (Reynolds et al. 1999) in human visual cortex (Reddy et al. 2009).

### ***Topic 2: Cue invariant, high-level object representations***

In chapter 3, using 3D structure-from-motion (SFM)- and luminance-defined face stimuli, we demonstrate that the amplitude of ERP components does not provide an optimal measure of categorical object perception. Specifically, we found that the ERP amplitude elicited by both static faces (N170) as well as by SFM faces (N250) was not cue invariant but rather highly dependent on visual stimulus fea-

tures within the cue dimension. This constitutes an “inverse problem” for category specificity and has implications for the interpretation of a large body of literature claiming a direct link between categorical perception and ERP peak amplitude (Bentin et al. 1996; Eimer 2000). Other approaches such as the face-inversion effect (Rossion et al. 1999, 2000) may be a better indicator for face-specificity since in this manipulation most low-level visual properties remain intact while our face-processing expertise is destroyed. However, while thanks to its high temporal resolution, ERP is an attractive method for studying many aspects of human behaviour, fMRI may constitute the more suitable tool for studying cue invariant object representations in the human visual cortex.

### ***Topic 3: 3D SFM-defined faces: a sensitive tool to dissociate normal from pathological functioning***

In chapter 4 we focused on the use of high-level complex SFM-defined objects to reveal subtle differences between normal and pathological functioning in amnesic Mild Cognitive Impairment (MCI) patients. Simple SFM stimuli have previously been indicated to provide a useful tool in addressing this question (Lemos et al. 2011), presumably due to their ability to probe integration mechanisms and dorsal-ventral stream cross-talk. The present study, by using SFM stimuli defining *faces*, yielded additional ventral stream-related information that is relevant in describing the specific deficits in MCI, and that so far has not received much attention.

Specifically, we found increased activation in MCI compared to healthy elderly control participants in the superior temporal lobe and several subcortical regions that have previously been shown to be involved in dorsal-ventral integration (Morel and Bullier 1990; Cherubini et al. 2010). Moreover, we observed aberrant patterns of face-related processing in FFA/OFA of MCI patients that may be related to an overrepresentation of low-level features in these face sensitive regions. Interestingly, based on their activation in FFA/OFA during our SFM paradigm, participants could be classified as members of the MCI or the age-matched control group with 88.2 % accuracy.

It is important to keep in mind that the present fMRI design coupled with a relevant cognitive task, proved successful in MCI patients, who, apart from mild memory complaints, displayed normal general cognition compared with age-matched controls. This shows that task-related fMRI may hold promise as a valuable functional biomarker e.g. for pharmacological treatment effects at an early stage of AD. However, at more advanced stages of AD, the various memory deficits and other symptoms may provide strong limitations on the application of complex task-related fMRI design. In the future, skill learning paradigms (e.g. Kim and Park 2010) or decoded fMRI neurofeedback (Shibata et al. 2011), in combination with pharmacological approaches may be challenging but promising approaches for reversing/sparing visual and other functions in MCI and preclinical AD.

The aberrant fMRI activation patterns during SFM face perception in MCI are likely due to the intricate interactions in expansive cortical networks that are necessary to compute face representations based on motion information. Defining

stimuli by multiple cues is another manner to trigger subtle interactions among different networks. If it holds true that wide-spread cortical interactions are impaired in MCI, it might be interesting to assess the application of cue combination paradigms in this patient group. Visual deficits in MCI might not be limited to weakening of (some) single cue representations (e.g. motion (Thiyagesh et al. 2009)) but might also show a lack of cooperative integration for cue combinations for which such integration is expected (e.g. motion and disparity (Johnston et al. 1994)).

## **In conclusion**

We found unexpected evidence for the lack of shading-motion cue integration, and instead found suppressive competition (a finding that may explain the conspicuous absence of shading-motion cue combination studies in the literature). We found an unexpectedly large dependency on visual stimulus features of ERP peak amplitudes that are believed to signal category-specific responses. And finally, we found an unexpectedly strong difference between MCI and control participants in cortical activation patterns induced by SFM-defined faces. Some unexpected findings are harder to publish than others, but in the end when that is done, the research performed here could be anticipated to have some theoretical impact (shading-motion cue combinations may be better framed in a Bayesian context than in a MLE framework), some methodological impact (a warning about the interpretation of category-specific ERPs), and some clinical impact (a possibility that task-related fMRI might aid early diagnosis of pre-clinical AD).

## References

- Bentin S, Allison T, Puce A, Perez E, McCarthy G. 1996. Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*. 8:551–565.
- Bulthoff H, Mallot H. 1987. Interaction of different modules in depth perception: Stereo and shading. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Technical report. 965.
- Cherubini A, Péran P, Spoletini I, Di Paola M, Di Iulio F, Hagberg GE, Sancesario G, Gianni W, Bossù P, Caltagirone C, Sabatini U, Spalletta G. 2010. Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients. *J Alzheimers dis*. 19:1273–1282.
- Eimer M. 2000. The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport*. 11:2319–2324.
- Johnston EB, Cumming BG, Landy MS. 1994. Integration of stereopsis and motion shape cues. *Vision Res*. 34:2259–2275.
- Kalva H, Christodoulou L, Mayron L, Marques O, Furht B. 2006. Challenges and opportunities in video coding for 3D TV. In: *IEEE International Conference on Multimedia and Expo*. p. 1689–1692.
- Kawai T, Shibata T, Inoue T, Sakaguchi Y; Okabe K, et al. 2002. Development of software for editing of stereoscopic 3D movies. *Proc. SPIE 4660, Stereoscopic Displays and Virtual Reality Systems IX*, 58
- Kim N, Park J-H. 2010. Learning to perceive structure from motion and neural plasticity in patients with Alzheimer's disease. *Neuropsychologia*. 48:1464–1471.
- Knill D. 2007. Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J Vis*. 7:1–24.
- Lemos R, Figueiredo P, Santana I, Simões MR, Castelo-Branco M. 2011. Temporal Integration of 3D Coherent Motion Cues Defining Visual Objects of Unknown Orientation is Impaired in Amnesic Mild Cognitive Impairment and Alzheimer's Disease. *J Alzheimers dis*. 27:1–12.
- Lovell P, Bloj M. 2012. Optimal integration of shading and binocular disparity for depth perception. *J Vis*. 12:1–18.
- Morel A, Bullier J. 1990. Anatomical segregation of two cortical visual pathways in the macaque monkey. *Vis Neurosci*. 4:555–578.
- Orban G. 2011. The Extraction of 3D Shape in the Visual System of Human and Nonhuman Primates. *Annu Rev Neurosci*. 34:361–388.
- Preston TJ, Li S, Kourtzi Z, Welchman AE. 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J Neurosci*. 28:11315–11327.

- Reddy L, Kanwisher N, Van Rullen R. 2009. Attention and biased competition in multi-voxel object representations. *P Natl Acad Sci USA*. 106:21447–21452.
- Reynolds JH, Chelazzi L, Desimone R. 1999. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci*. 19:1736–1753.
- Rossion B, Delvenne JF, Debatisse D, Goffaux V, Bruyer R, Crommelinck M, Guérit JM. 1999. Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol Psychol*. 50:173–189.
- Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, Crommelinck M. 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*. 11:69–74.
- Shibata K, Watanabe T, Sasaki Y, Kawato M. 2011. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*. 334:1413–1415.
- Thiyagesh SN, Farrow TFD, Parks RW, Accosta-Mesa H, Young C, Wilkinson ID, Hunter MD, Woodruff PWR. 2009. The neural basis of visuospatial perception in Alzheimer’s disease and healthy elderly comparison subjects: an fMRI study. *Psychiat Res*. 172:109–116.

## English summary

Our visual system enables us to perceive the world in 3D by using a number of monocular (shading, texture and motion) as well as binocular visual depth cues (disparity). The body of work presented in this thesis focused on the recovery of 3D structure from shading and motion visual image cues as well as their underlying neural mechanisms in human observers. Specifically, we addressed 4 main research questions pertaining to 1) curvature estimation from single as well as combined shading and motion visual cues, 2) the neural mechanisms underlying curvature estimation from single- and combined shading and motion cues, 3) the applicability of ERP amplitude measurements as a method for deriving high-level object-specificity from shading- and motion-defined objects and 4) the specific deficits in perceiving 3D motion-defined objects in mild cognitive impairment (MCI) patients.

**Chapter 1** was motivated by the question of how low level 3D curvature properties of simple objects are perceived when defined by either single shading, single motion or by combined shading and motion visual image cues. We probed curvature estimation using a psychophysical curvature judgment task, comprising two single cue conditions and two combined cue conditions, in which shading and motion cues were either consistently or inconsistently combined. Psychophysical results revealed that contrary to the prediction of Maximum Likelihood Estimation (MLE), shading and motion cues did not combine in a statistically optimal fashion. Instead, reliability of curvature judgments in combined compared to single cue conditions was *reduced*, suggesting a form of suppressive competition between these specific cues. Together with the large dominance of the shading over the motion cue in the consistent combined condition which could not be explained by differences in single cue reliabilities, our data suggest that MLE is not applicable to any arbitrary choice of cue pairs. Instead, several ecological factors, the role of experience and prior knowledge about the world may put constraints on cue integration, determining which cues do and which do not integrate.

In a follow-up study we intended to extend the findings presented in Chapter 1 and continued the investigation of shading and motion cue combination by employing a curvature discrimination task coupled with event-related fMRI (**Chapter 2**). While the psychophysical results in this study supported the previous findings from Chapter 1, showing a lack of cooperative integration and instead a weighted averaging of shading and motion cues in combined conditions, their underlying neural correlates were difficult to interpret. The lack of a parametric relationship between stimulus curvature (defined by either single or combined cues) and fMRI signal in the group average, suggests that curvature-related information may not be represented by spatial averaging of neuronal populations but rather by multivariate patterns of BOLD signals. Although in the present analysis we found that fMRI-based curvature sensitivity in combined cue conditions could be predicted from a combination of fMRI-based curvature sensitivities in single cue conditions in a right IPS region, the implication of these findings concerning low-

level curvature coding and shading-motion cue combination are difficult to understand at present and deserve further analysis/investigation.

While in chapter 1 and 2 we studied low-level curvature properties in relatively simple 3D shapes, the work presented in chapter 3 and 4 concerns the study of high-level object vision. With a large body of literature claiming a direct link between categorical perception and the amplitude of ERP components - the most famous example being the face-specific N170 component - we addressed in **Chapter 3** the *cue invariance* of 'object specific' ERPs. By using 3D structure-from-motion (SFM)- and luminance-defined face stimuli, we found that the ERP amplitude elicited by both static faces (N170) as well as by SFM faces (N250) was not cue invariant and instead highly dependent on visual stimulus features within the cue dimension. This implies an "inverse problem" for category specificity and suggests that the amplitude of ERP components may not provide a robust measure of object perception.

Thanks to advancements in the knowledge of normal perceptual processing, we may also begin to better understand perceptual losses due to ageing, disease or brain damage. In **Chapter 4** we focused on high-level object vision in Mild Cognitive Impairment (MCI) patients, a medical condition that shows a high risk for conversion to Alzheimer's disease (AD). To reveal subtle differences between normal and pathological visual functioning, we used complex SFM-defined object stimuli which provide a valuable tool to study integrative function and dorsal-ventral cross talk. We observed aberrant patterns of face-related processing in the face-sensitive temporal lobe regions FFA/OFA of MCI patients that may be related to an overrepresentation of low-level features. Interestingly, based on their activation in FFA/OFA during our SFM paradigm, participants could be classified as members of the MCI or the age-matched control group with 88.2 % accuracy. Moreover, we found increased activation in MCI compared to healthy elderly control participants in the superior temporal lobe and several subcortical regions that have previously been shown to be involved in dorsal-ventral integration. These results suggest that ventral stream information, in addition to dorsal-stream deficits, is relevant in describing the specific deficits in MCI. Furthermore we concluded that task-related fMRI may hold promise as a valuable functional biomarker, e.g., for pharmacological treatment effects at an early stage of AD.

Overall, the work included in this dissertation contributes to the understanding of low- and high-level 3D vision from different visual depth cues, the use of ERP peak amplitudes for investigating high-level object vision and the specific visual deficits related to 3D vision in MCI patients. Intricate interactions between ventral and dorsal processing pathways (or restrictions thereof), in both lower- and higher-order visual areas likely play a significant role in explaining the present findings and should be addressed by integrating various fields of research in the future.

## Samenvatting

Ons visuele systeem stelt ons in staat de wereld in drie dimensies (3D) waar te nemen, door gebruik te maken van een aantal monoculaire (schaduw, textuur en beweging) evenals binoculaire diepte cues (dispariteit). Het onderzoek in deze thesis focust op hoe mensen 3D structuur afleiden uit visuele cues zoals schaduw en beweging en uit de kenmerken van de onderliggende neurale mechanismes.

We behandelen 4 belangrijke onderzoeksvragen namelijk 1) de schatting van curvatuur uit één enkele -of gecombineerde schaduw en visuele beweging prikkels, 2) de neurale mechanismes onderliggend aan inschatting van curvatuur uit één enkele of uit gecombineerde schaduw- en visuele beweging prikkels, 3) de toepasbaarheid van EEG als een methode voor het afleiden van hogere orde objectspecificiteit aan de hand van door schaduw en beweging gedefinieerde objecten en 4) de specifieke tekortkomingen in de waarneming van drie dimensionale –door beweging gedefinieerde- objecten bij patiënten met mild cognitive impairment (MCI).

**Hoofdstuk 1** is ontstaan naar aanleiding van de vraag over hoe lagere orde 3D curvatuureigenschappen van eenvoudige objecten worden waargenomen wanneer deze worden gedefinieerd door ofwel enkel schaduw, enkel beweging of een combinatie van schaduw en bewegingcues.

Wij onderzochten de inschatting van curvatuur door middel van een psychofysische taak waarbij proefpersonen de mate van curvatuur in moesten schatten. Schaduw en beweging cues werden in de experimentele condities zowel op consistente wijze als inconsistente wijze gecombineerd. De resultaten uit dit experiment lieten zien dat in tegenstelling tot de voorspellingen van ‘Maximum Likelihood Estimation’ (MLE), schaduw en beweging cues niet combineerbaar waren in een statistisch optimale wijze. In tegenstelling, de betrouwbaarheid van de inschatting van de mate van curvatuur werd *verlaagd* door de combinatie van cues in vergelijking met zelfstandige cues. Dit duidt op een mogelijke competitie tussen deze specifieke cues. Dit samen met de grote dominantie van de schaduwcue, in vergelijking met de bewegingscue, in de ‘consistent gecombineerde’ conditie (welk niet verklaard kan worden door verschillen in de betrouwbaarheid van de zelfstandige cues) suggereert dat MLE niet van toepassing is op een willekeurige combinatie van cue soorten. Het lijkt erop dat ecologische factoren, de rol van ervaring en voorkennis, de mate van cue integratie beperkt en zo bepaalt welke cues wel en welke cues niet te integreren zijn.

In een vervolgstudie wilden we verder borduren op de bevindingen uit Hoofdstuk 1 en vervolgden het onderzoek van schaduw -en bewegingscue combinatie door middel van een fMRI studie waarin proefpersonen een kromming discriminatie taak uitvoerden (**Hoofdstuk 2**).

De psychofysische resultaten bevestigden de eerdere bevindingen uit Hoofdstuk 1, namelijk het uitblijven van coöperatieve integratie maar een resulterend gewogen gemiddelde van schaduw en beweging cues in de

'gecombineerde' condities. De resultaten bemoeilijken de interpretatie van onderliggende neurale mechanismen. Het ontbreken van een parametrische relatie tussen de curvatuur van de stimulus (gedefinieerd door ofwel zelfstandige ofwel een combinatie van cues) en de grootte van het fMRI signaal suggereert dat kromming gerelateerde informatie wellicht niet optimaal wordt gerepresenteerd in het ruimtelijk middelen van de respons van neurale populaties maar wellicht beter zichtbaar is in multivariate patronen van BOLD activatie. Alhoewel we in de huidige analyse van het rechter IPS gebied vonden dat de sensitiviteit van fMRI voor curvatuur bij stimuli met gecombineerde cues voorspeld kon worden uit de sensitiviteit van fMRI voor curvatuur bij stimuli met enkelvoudige cues. De implicatie hiervan met betrekking tot lagere orde curvatuur codering en schaduw-beweging cue combinatie is op dit moment nog moeilijk te interpreteren en vereist verdere analyse en verder onderzoek.

Waar we in hoofdstuk 1 en 2 lagere orde curvatuur eigenschappen bestudeerden in relatief eenvoudige 3D vormen, bestuderen we in hoofdstuk 3 en 4 de hogere orde object waarneming. Aangezien een groot deel van de literatuur duidt op een directe link tussen categorische perceptie en ERP componenten – het meest bekende voorbeeld is het gezicht-specifieke N170 component – bestudeerden we in **Hoofdstuk 3** het fenomeen 'cue-invariance' van 'objectspecifieke' ERP's. Gebruik makend van 3D structure-from-motion (SFM) -en luminantie gedefinieerde gezicht stimuli, vonden we dat de amplitude van het ERP signaal, opgewekt door statische gezichten (N170) alsmede door SFM gezichten (N250), niet cue invariant maar juist sterk afhankelijk was van visuele stimulus eigenschappen binnen de cue dimensie. Dit impliceert een inversieprobleem voor categorische specificiteit en suggereert dat de amplitude van ERP componenten wellicht niet de meest robuuste maat is voor object perceptie.

Dankzij grote vooruitgang in de kennis over gezonde perceptuele verwerking kunnen we wellicht een begin maken met het beter begrijpen van perceptuele problemen veroorzaakt door veroudering, ziekte of hersenbeschadiging. In **Hoofdstuk 4** concentreerden we ons op hogere orde object waarneming in Mild Cognitive Impairment (MCI) patiënten, een medische aandoening waarbij patiënten een verhoogd risico lopen op het ontwikkelen van de ziekte van Alzheimer. Om de subtiele verschillen tussen gezond en pathologisch visueel functioneren te onthullen, maakten we gebruik van complexe SFM-gedefinieerde objectstimuli welke uitermate geschikt zijn voor het bestuderen van het vermogen tot integreren en dorsaal-ventrale intercommunicatie. De resultaten wezen op abnormale patronen van gezicht gerelateerde verwerking bij MCI patiënten in de temporale hersengebieden die bekend staan om hun gevoeligheid voor gezichten FFA/OFA. Deze abnormale patronen zijn wellicht gerelateerd aan een over-representatie van lagere orde eigenschappen van de stimuli. Interessant genoeg maakt dit patroon van activatie in FFA/OFA tijdens het SFM paradigma het mogelijk om proefpersonen met 88.2% nauwkeurigheid te classificeren als MCI patiënt of lid van de controlegroep (gezonde ouderen). Voorts vonden we een verhoogde activatie bij MCI patiënten, in vergelijking met de controlegroep, in de superieure temporaal kwab en een aantal subcorticale gebieden waarvan eerder is aangetoond dat deze

zijn betrokken bij de dorsaal-ventrale integratie. Deze resultaten suggereren dat het patroon van informatie binnen de ventrale stroom, samen met dorsale stroom gebreken, een belangrijk deel uitmaakt van MCI gerelateerde symptomen. Wij concludeerden dat taak gerelateerde fMRI veelbelovend kan zijn als een vorm van 'bio-marker' tijdens het vroege stadium van de ziekte van Alzheimer.

Het werk beschreven in deze dissertatie draagt bij aan het begrip van lagere - en hogere orde 3D waarneming aan de hand van verschillende visuele diepte cues alsook aan het gebruik van ERP piek amplitudes voor het onderzoeken van high-level object waarneming en de specifieke visuele gebreken gerelateerd aan 3D visuele waarneming bij MCI patiënten. Complexe interacties tussen ventraal en dorsale verwerkingspaden (of de beperktheid daarvan), in zowel de lagere - alsmede de hogere visuele gebieden, zullen hoogstwaarschijnlijk een belangrijke rol spelen in de interpretatie van de huidige resultaten en zouden verder moeten worden geadresseerd door toekomstig multidisciplinair onderzoek.



## Acknowledgements

Knowing that the acknowledgment section is among the first things being read in this scientific publication (and rightly so, it gives credit to the people that helped to make it all happen!), I won't miss my chance to engage in non-scientific chitchat to the fullest.

First and foremost, I would like to thank my supervisors Miguel Castelo-Branco and Peter De Weerd. Miguel, I'm very glad to have been given the opportunity to spend the first two years of my PhD at IBILI working with such a knowledgeable and enthusiastic supervisor. Your wide range of scientific interests and the variety of projects you are working on has been a very motivating experience and made me look at my own work from different perspectives. Peter, from your diplomacy and wisdom and by being the strictest reviewer of our papers yourself I have learned a lot which I am very grateful for. Thank you so much for your thoughtful, thorough, and friendly advice and support throughout the past few years.

My second acknowledgment must go to my lab friends at IBILI. First of all, Inês A., thanks for being the nicest office mate and smilingly coping with my occasional Matlab-, Statistical-, Administration- or general scientific crises. I miss our chats and I'm looking forward to run a half-marathon together soon. Let's visualize that. Also, Inês B. and Inês V., (to complete the Inês Trio), thank you for fun lunches, a cool trip to Canada and for being scared when I pretended to have heard a bear knocking on the 'door' of our tent. Capriola, Zézinho, Arroz de Rebola. Es fixe! Tantas saudades de resting at the movies with you, saying 'sauuuude', Sundays lunches at your grandma's house and going to the beach at 5.30 pm to make sure the sun has already set. Nobody appreciates my singing (humming) as much as you do! Filipa, Raquel, Joana Sampaio, Marieke, Castelhana and Ana Pina, I very much enjoyed our nice lunches, cafés and our cultural/scientific discussions. Thank you also to Carlos and Joaozinho for their help and advice with scanning.

Paso and Marta desayuno, thanks for your unstoppable Portuguese cultural and culinary wisdom, for letting me throw bread off a church tower in Aveiro, beach days and for inviting me to your wedding. It was wonderful! Marcozinho, os jantares no Arco Bar contigo foram inesquecíveis! A sério? Miúdos de leitão? Eu sempre prefiro um Cha de Saquentinho. Só em Vizzzeu tomo Caipirinhas e de manha um pouco de mel com figos (caídos do arvore).

Maria, Cindy, Alex e Pablo da Rua Doutor Dias Ferreira. Tenho muitas saudades da nossa dança da casa, de sairmos juntos à noite, dos jantares e das cookies. Vocês são os melhores colegas da casa do mundo!

Malta 'Coimbra noite 2008/2009'. Vocês são *legendary*! Adorei as sexta-feiras em que fomos a um sítio diferente, tipo ao Tropical. Joana S., obrigada para me levares às primeiras festas em Coimbra com esta malta louca, pelos jantares no Tapas, para me ensinares a comer Gambas e para falar o alemão mais fixe do mundo. Celulazinhas! Manu, Giana, Patrick, Ivan, Tiago, Vitorino, Marco Matosão, Joana Costa; tenho saudades vossas, das Super Bocks no Tropi, na Associação e no Shots bar e também (mas só um pouco) das noites no el Divino! Passar tempo com vocês foi a melhor compensação de uma vida científica dura.

To my outside-of-science friends from somewhere along the way.

My dear flatmate from Rua Augusta Rocha Jean François. Thanks for volunteering to participate in my experiments at the lab, for approx. 100 city tours of Brussels, the best Belgian fries ever and for keeping in touch. It's always a pleasure to spend time with a pro-surfer called Joao-Francisco.

Simon, your confidence and support often helped to make me smile instead of being stressed and have been invaluable during the early stages of my PhD. Thanks for having been there, lending your best Matlab support, fun times in Peniche, London, Scotland etc and for keeping in touch since we first met in SA.

Miriam. Wie gut, dass wir damals im Methoden & Techniken Kurs endlich mal miteinander gesprochen haben! Zur Uni fahren, für die Toetsen lernen, Party machen und Barcelona- und Australienurlaube hätten ohne Dich und unsere guten Gespräche nur halb so viel Spass gemacht!

Melly. Witze erzählen im Heustall, unsinnige Telefonanrufe bei meinem Musiklehrer (lange her), im Ka zur Uni fahren, Italienurlaube und zuletzt Postkarten aus Australien. Danke für Deine Freundschaft in den letzten 19 (!) Jahren, Deine treue Unterstützung in allen Lebenslagen, Dein Verständnis für meinen PhD Stress, Deinen Pragmatismus, klaren Blick, sehr sehr viele lustige Momente und dass wir beide immer automatisch zu spät kommen wenn wir uns treffen!

Jo, Julia, Laura, Lars M., Markus und Vera: Ich bin froh, dass ich Euch habe und Ihr immer wieder für einen spassigen Ausgleich zur wissenschaftlichen Welt sorgt.

To my lovely Capetonian friends Gailyn and Craig. You were there when I received the email saying that I had won the PhD scholarship in November 2007 and you were there when I received the email saying that my thesis got approved in March 2013. So I suppose it's fair to say that you were involved from the beginning until the end. ☺ Despite not having seen you in between and living 14.242 km away from Hout Bay, I am very happy we've stayed in touch. Thanks for being so lovely and down to earth and making me feel at home and like I never left whenever I am back in South Africa.

Ollie, thanks for taking me sailing, sharing many beautiful beaches, carrying my new couch, lovely dinners at Café Sjiem and being as spontaneous as you are. I loved all our adventures.

Dear CN friends in Maastricht, thank you for creating such a positive and stimulating working environment! I think my real introduction to CN people & life was HBM 2010 in Barcelona. The spirit lasted until long after the conference. A scientific motivation boost along with fun times, the “group activity thing”, great discussions and the most extreme tiredness upon returning to Maastricht. Thanks all, that was fantastic!

To my office mates.

Lars Riecke. During the 1.5 years in the UNS40 office with you we had the perfect balance between concentrated silence (okay, interrupted by regular intervals of your teapot making some strange sizzling sound) and very enjoyable and funny chats. Thanks for all the questions that I was allowed to throw at you and that you would patiently answer and for all the good stories!

Sanae and Valentin. When an Anthropologist, a Physicist and a Psychologist share an office and they all do Neuroscience, the combination can only turn out to be great! Thank you for the help, sharing thoughts, the laughs, good conversations and a very enjoyable office atmosphere.

To my Paranymfen.

Just a few steps away from the ‘old’ office: Marin. Thanks so much for your help & advice in all PhD matters over a cup of tea, for proofreading my introduction, for understanding my PhD stress and being happy for me when the thesis was finally done. I’m looking forward to more climbing, bouldering and shopping sessions.

Just a few more steps away from the office: Anke. Thanks for your friendship, encouraging words about non-existing results when they were most needed (both the results and the words), frequent lifts from/to Aachen as well as pre-movie/pre-PhD party dinners and catching up at my flat. See you and Sven soon on your (future) rooftop terrace in Cologne for BBQs.

Someone who randomly drops by the office: Felix! You’ve been a constant source of inspiration, excellent mental support and have always provided a very relaxed and clear view on all sorts of things. Besides that, your jokes are really funny (in English too). Thanks also to Gesa for all-you-can-eat cooking events, hiking and Sunday afternoon cycle tours.

Coffeelovers, 1992 plein, Saturday at 1pm. Marin is already there and orders a Boerenbroodje met oude kaas and a Coffeelovers Latte. We sit down and Lars shows up (‘I’ve just delivered some results. I’ll have a Cappuccino please.’). 30 minutes later and after having skyped with his friends back home, Matteo arrives. He orders a Ciabatta BTLTC. Same as Joao who has just come back from Portugal and is about to leave for xxxx. If you are lucky, Martin H. shows up as well and orders a softdrink. They say that simple routines contribute to more happiness in every day life. I couldn’t agree more. Having said this, Alex arrives (2.30 pm) and orders an Espresso (‘I was chilling at home for a bit’). Zondag or Take 5 tonight?

Special thanks to Lars and Alex for making excellent housemates and taking such good care of the only girl in the house. Thanks for cooking me dinners, watching movies, having sleep-inducing teas and watching most exciting boxing matches together.

Thanks also to Martin, Francesco, Nick and Giancarlo for accepting me into your prestigious lunch group when I first got back to Maastricht. It's been a pleasure to talk about cross-cultural differences and Matt Damon among other important things. And of course unforgotten: The consequences of a visit to an icecream shop at the Ramblas. Chapeau guys, that was hilarious!

Francesco; as promised I am also going to acknowledge your 10 years of patience with my impulsive driving, parties at the nightlife in my first year of Uni, snowboarding lessons, Aachen Christmas market visits and lifts to Miami airport (I'm *eternally* grateful for that).

Michelle, Tom, Job, Kamil, Marieke, Aline, Rosanne, Roberta, Anna G. Jan, Anna Z., Mario, Gojko, Mehrdad, Sanne. I've very much enjoyed every coffee/lunch break with you. Very grateful to have had such great colleagues and to have met so many nice people! I would also like to thank Bert and Michael Capalbo for valuable and funny scientific advice as well as rock climbing assistance. Bert, thanks to you I always felt like I was ahead of schedule!

Thanks also to Annemie, Christl and Riny for their competent, patient and good-humoured administrative support and especially for their help in reconciling Portuguese and Maastricht administration.

Maurice Hartmans, nobody makes email exchanges that concern printing and system access issues as extra-ordinary as you do. Thanks for the laughs and the occasional words of the day. Jordi, thank you for bringing my computer back to life and saving my hard drive on a black Wednesday approximately 1 month before the deadline to hand in my thesis. I am extremely grateful.

Liebe Betty, lieber Hans. Danke für ein immer offenes Ohr und eine immer offene Tür. Eure unkomplizierte und bodenständige Art hat mir so manches Mal geholfen die Dinge in der richtigen Relation zu setzen, die Wissenschaft zu vergessen oder auch mal herzlich darüber zu lachen.

Ausserdem möchte ich Dirk, Barbara, Theresa, Valentin, Heike, Klaus, Sassi und Natalie danken für Euer Verständnis und Euren humorvollen Umgang mit dem Thema Doktorarbeit. Wenn Ihr nächstes Jahr an Weihnachten fragt, dann ist die These auch tatsächlich fertig!

Von ganzem Herzen danke ich auch meiner Mutter. Dein Vertrauen in mich, Deine Zuversicht und Deine uneingeschränkte Unterstützung auf all meinen Wegen sind mir unendlich wertvoll. Danke für kluge und praktische Ratschläge, Deine Energie, Stärke und Fröhlichkeit und dass Du immer für mich da bist.

## Curriculum Vitae

Britta Graewe was born on June 23<sup>rd</sup> 1983 in Aachen, Germany. She attended the Couven College where she completed her secondary education in 2002. In 2003 she commenced to study Psychology at Maastricht University. During an exchange program with the University of California, Santa Barbara in 2006 she specialized in Biological Psychology and, upon returning to Maastricht, enrolled in the Cognitive Neuroscience Master program. In the scope of her research internship at the University of Coimbra, Portugal she investigated visual motion perception under the supervision of Prof. Elia Formisano and Prof. Miguel Castelo-Branco. She graduated cum laude from Maastricht University In 2007. In 2008, Britta received an FCT PhD scholarship for a joint project between the Faculty of Medicine, University of Coimbra and the Cognitive Neuroscience Department at Maastricht University. Supervised by Prof. Miguel Castelo-Branco and Prof. Peter De Weerd she investigated the neural mechanisms of 3D vision in healthy participants and visually impaired patients.