Object Recognition by Integration of Information Across the Dorsal and Ventral Visual Pathways

Reza Farivar-Mohseni

Department of Psychology
McGill University
Montréal, Quebec, Canada

July 2008

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

ABSTRACT

The brain decomposes visual information into its form and motion components and processes the two aspects largely independently by way of anatomically distinct pathways that originate early in the visual system and continue ventrally to the occipito-temporal visual areas and dorsally to the occipito-parietal visual areas, respectively. Certain cues of shape, such as 3-D structure-from-motion (SFM), appear to be computed exclusively by dorsal-stream mechanisms, yet these cues can describe complex objects whose recognition depend on mechanisms in the ventral stream. This dissertation discusses theoretical means by which dorsally-computed 3-D cues may provide input to ventral stream object recognition mechanisms. Psychophysical and neuropsychological data presented here suggest that 3-D SFM cues do indeed empower complex object recognition, and recognition of shapes defined by 3-D SFM do likely require integration of information across the two pathways. Additionally, neuropsychological data are presented for a dissociation of 3-D SFM processing from 2-D form-from-motion processing. Finally, utilizing functional imaging (FMRI), data are presented to suggest that SFM-defined objects do not engage category-selective areas in the human brain in the same manner as photographs of those objects do. Together these results suggest that visual object recognition may be subserved by mechanisms distributed between the two pathways.

Résumé

Le cerveau décompose l'informations visuelle en ses composants de forme et de mouvement, et les traite de manière indépendante par deux voies anatomiques distinctes—l’information ayant attrait au mouvement et à la relation spatiale par la voie dorsale qui se termine dans le lobe pariétal et l’information ayant attrait à la forme par la voie ventrale qui se termine dans le cortex inférotemporal. Certaines informations de profondeur, tel que la structure-par-mouvement 3-D (SPM), sont presque entièrement analysées par la voie dorsale; toutefois, les objets décris par la SPM sont aussi reconnus par les voies ventrales. Cette thèse débute par une discussion théorique décrivant la manière dont l’information de profondeur calculée par la voie dorsale peut contribuer aux machinismes de reconnaissance des objets (voie ventrale). Les résultats des expériences psychophysiques et neuropsychologiques indiquent que l’information de SPM peut permettre la reconnaissance des objets complexes, même des visages peu familiers, et cela peut constituer un case d’intégration entre les deux voies indépendantes. De plus, les résultats des expériences neuropsychologiques présentées suggèrent que la perception de forme-par-mouvement 2-D est dissociable de celle de structure par mouvement 3-D. Finalement, par le biais d'imagerie par résonance magnétique fonctionnelle, nous avons démontré que les objets décris par SPM n’activent pas le même méchanisme cérébral que des photos de ces mêmes objets. Ensemble, les résultats présentés ci-après suggèrent que la reconnaissance des objets visuels peut être distribuée entre les deux voies visuelles.

ACKNOWLEDGEMENTS

I am thankful to many people for their guidance, encouragement, and support throughout my PhD program. First and foremost, I wish to thank my advisor, Prof. Avi Chaudhuri, who took me in and provided me with an open atmosphere to develop. My colleagues and friends in the CVL deserve many thanks, particularly Carmelo Milo, Karen Borrmann, and Jae-Jin Ryu. I also with to thank members of the McGill Vision Research, Prof. Curtis Baker, Athena Buckthought, and Benjamin Thompson, for their feedback and comments on earlier drafts of this dissertation.

Some of this work would not have been possible without the help of collaborators. I wish to thank Prof. Olaf Blanke for providing me with the opportunity of testing unique patients and Michela Adriani and Manuel Mercier in the Blanke lab for their help with various aspects of the projects. Special thanks to Prof. Guy Orban and Prof. Ruffin Vogels and Simon Clavigny for helpful comments and encouragement on Chapter 2. I am grateful also to Prof. Bruno Rossion for allowing me to use his facilities to test Patient P.S. and for his comments and input on some of the experiments.

I am indebted to my wonderful friends for their enduring support and encouragement. Particular thanks to Penelope Kostopoulous for butt-kicking, Marie-Claire Albanese for tuna pasta and philosophical discourse, Roger Laforce and Marjorie Rabiau for making me fat and being wonderful, David Paul for listening and challenging, Craig and Anna Aaen-Stockdale for all the babysitting and wonderful nights, and Jurgen Germann for conversations and thoughts.

I wish to acknowledge and thank my grandparents, Baba Hussein and Azamjoon, without whom my love for learning would have never flourished. I am fortunate to have had such grandparents and will take their loving memory for all my life. I also wish to extend special thanks to my uncle (almost like a second father), Daryoush Diba, for keeping me working, whether sweating on his garden or over books in the library.

Particulièrement, je voudrais remercier mes beaux parents, Ilona et Karel Gazurek, pour tout le soutien qu'ils m'ont donné pendant les mois que j'ai passé chez eux, pendant ma collaboration avec Olaf Blanke à Genève. Si je suis arrivé à parler en française un peu mieux, c'est a cause de Madame La Chose et Il Signore.

Not one aspect of this would have been possible without the many sacrifices made and the abundance of love and encouragement showered on me by my mother. I hope that I could live up to all that you have done for me. I wish to thank my father for his constant encouragement and his unwavering belief in me, and my brother and sister for their constant supply of love and challenges.

Finally and most importantly, I am most grateful to the love of my life, my soul mate and my loving partner, Eva. She is the light of my life on dark days, denying my self-doubt and encouraging me throughout all the obstacles that fell in my way. I am grateful to her for believing in me, for trusting me and encouraging me to push on when my motivation was fading. Eva, thank you. You are more than anything I could dream of. I love you more than I could ever express.

To the memory of Baba Hussein and Azamjoon

To my mother and father

To Eva and Bastian, for everything.

ORIGINAL CONTRIBUTIONS TO KNOWLEDGE

This doctoral thesis presents a number of original contributions about the nature of cortical organization, particularly pertaining to the dorsal and ventral visual pathways, as well as on the integration of information across the two pathways.

Chapter 2 presents a novel synthesis of current models of object recognition and cortical representation of object shape information.

Chapter 3 presents psychophysical results from naïve subjects showing that 3-D structure-from-motion (SFM) cues can empower complex object recognition, such as the recognition of unfamiliar faces. Additionally, neuropsychological evidence is provided to support a role for dorsal-ventral integration in the recognition of motion-defined faces. The results support the view that object recognition is distributed across the dorsal and ventral pathways.

Chapter 4 represents data collected from a number of neurological patients who together exhibit a functional dissociation between 2-D form-from-motion deficits and 3-D structure-from-motion deficits. These results suggest that while motion may inform of shape, there may be multiple dissociable mechanisms of form or structure from motion.

Finally, Chapter 5 presents results from a functional magnetic resonance imaging (FMRI) study investigating the cortical basis for 3-D SFM face recognition. The results suggest that these stimuli do engage category-specific regions in the ventral pathway, but not the Fusiform Face Area. These results place a greater role for the occipital face area in the processing of 3-D information of a face.

CONTRIBUTION OF AUTHORS

The contribution of the authors to the manuscripts on which this thesis is based is as follows. Reza Farivar developed the major parts of methodology, designed and conducted the experiments, and wrote the bulk of the manuscripts. The supervisor, Dr. Avi Chaudhuri, and collaborators, Drs. Michael Petrides and Olaf Blanke, provided guidance in the different fields of functional imaging and neuropsychological testing.

PREFACE

This is a manuscript-style thesis that provides new insights on the integration of information across the dorsal and ventral visual pathways.

The dissertation is based on the following manuscripts:

Farivar, R. (2008). Dorsal-ventral integration in object recognition. *Brain Research Reviews.* Conditionally accepted.

Farivar. R., Blanke, O., and Chaudhuri, A. (2008). Dorsal-Ventral Integration in the Recognition of Motion-Defined Unfamiliar Faces. *Journal of Neuroscience.* Re-submitted.

Farivar. R. Chaudhuri, A., Blanke, O. (2008). 2-D Form-from-Motion Deficit with Intact 3-D Structure-from-Motion Perception. In Preparation.

Farivar. R., Petrides. M., and Chaudhuri, A. (2008). Cortical representation of complex objects defined by 3-D structure-from-motion. *Brain Research*. Submitted.

These papers appear in Chapters 2 through 5, respectively.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

List of Abbreviations

2-D, 3-D	Two-dimensional, three-dimensional
AIP	Anterior intrraparietal area
CIP	Caudal intraparietal area
EEG	Electroencephalography
EPI	Echo-planar imaging
FFA	Fusiform face area
FFM	Form-from-motion
FMRI	Functional magnetic resonance imaging
FOV	Field-of-view
FWHM	Full-width-half-maximum
GE	Gradient echo
hMT+	Human homolog of macaque MT
HRF	Hemodynamic response function
IPS	Intraparietal sulcus
LCD	Liquid crystal display
LIP	Lateral intraparietal area
LO	Lateral occipital area
LOC	Lateral occipital complex
MEG	Magnetoencephalography

MRI	Magnetic resonance imaging
MST	Satellite of area MT
MT	Middle temporal area, V5
OFA	Occipital face area
PC	Percent correct
ROI	Region of interest
SFM	Structure-from-motion
STS	Superior Temporal Sulcus
TE	Time to echo
TMS	Transcranial magnetic stimulation
TR	Repetition time
V1	Visual area 1
V2	Visual area 2
V3	Visual area 3
V4	Visual area 4
V5	Visual area 5, MT

Everything flows. Nothing stands still.

Heraclitus of Ephesus

Chapter 1

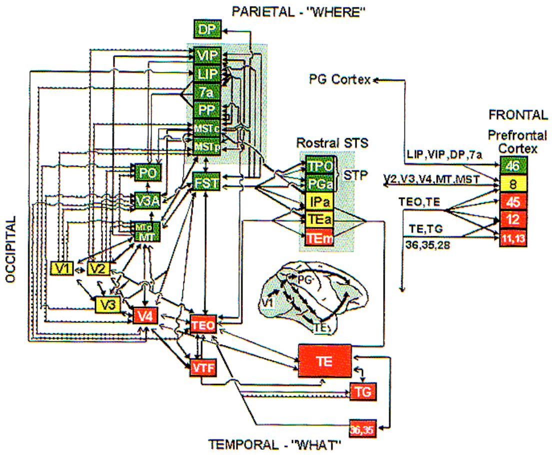
INTRODUCTION

Since the early days of the neural sciences our ideas of cortical organization have been shaped around a view of the brain as a segmented organ. One of the most influential ideas in cortical representation of vision was proposed by Ungerleider and Mishkin (1982). Based on the pattern of object recognition deficits following temporal lesions (Brown and Shafer, 1888; Klüver & Bucy, 1937; Mishkin, 1954) and spatial disorientation following posterior parietal lesions (Holmes and Horax, 1919; Pohl, 1973), Ungerlieder and Mishkin proposed two cortical processing streams, one coursing dorsally from striate cortex to the posterior parietal regions and subserving spatial vision, and another coursing towards the inferior temporal cortex and concerned with object or identity recognition. The seminal 1982 chapter has now been cited over 2000 times and has indirectly influenced a plethora of research while also implicitly guiding our ideas of cortical structure and function, both in vision and in other sensory modalities. A summary of some evidence supporting the dual-pathway model is provided in the following pages. This is followed by a discussion of some outstanding issues concerning the overlap of function between the two streams.

Neuroanatomical Basis for Separate Pathways

It should be noted that all our knowledge of the primate visual system is derived from connectivity studies in the monkey brain (mainly macaque), and in fact almost all single-unit electrophysiology findings are also obtained from the monkey brain.

The division of the two pathways occurs early. The afferents from thin stripes and interstripe regions of V2—anatomical subdivisions of V2 based on staining of cytochrome oxidase—are forwarded to V4 (DeYoe & Van Essen, 1985; Shipp & Zeki, 1985), while those of the thick stripes provide input to V3 and together with V1 to MT (Roe & Ts'o, 1995; Shipp & Zeki, 1989). Baizer et al. (1991) supported the dual-pathway model in a direct assessment of neuroanatomical connectivity. They found that retrograde injection of tracers into a posterior parietal area (LIP) resulted in staining in areas MT/V5, V3, V2 and V1, while barely staining middle temporal regions. In contrast, retrograde tracer injections into the inferior temporal gyrus (IT) resulted in staining in posterior IT (TEO), V4, V3, V2, and V1, with little staining in posterior parietal regions or V5/MT. The figure on the left summarizes the connectivity of the macaque visual system.



While the notion of separate anatomical pathways has withstood the test of time, it must be noted that the two pathways are interconnected and it is at times difficult to decide the organization of the system as its organization is largely underdetermined (Hilgetag, O'Neill, & Young, 1996). By assuming certain rules, it is possible to derive specific patterns. For example, Young (1992) suggests strong segregation of the two pathways by clustering areas together based on the density of their connections—if two areas have strong connections (feedforward and feedback) then they were deemed "closer" in the hierarchy. Based on this assumption, the largely accepted view of the parallel hierarchies in vision appears to hold.

Functional properties attributed to each stream

Dorsal Stream

Neural selectivity

A particularly unique property of dorsal stream areas is their sensitivity to visual motion. While direction selectivity is exhibited in "early" visual areas, such as V1 (Hubel & Wiesel, 1968), V2 (Burkhalter & Van Essen, 1986), and V3 (Gegenfurtner, Kiper, & Levitt, 1997; Zeki, 1978), the largest number of cells exhibiting motion selectivity are found in V5/MT (Albright, 1984; Albright, Desimone, & Gross, 1984; Zeki, 1978). Furthermore, V5/MT cell firing correlates with the reported perception of the direction of rotation of a transparent cylinder defined by structure-from-motion (SFM; Grunewald, Bradley, & Andersen, 2002). However, it is unclear whether the SFM selectivity observed by these authors is due to perception of depth from SFM or sensitivity to motion transparency without any depth perception. It is known that response of V5/MT cells can distinguish motion signals in motion transparent displays (Recanzone, Wurtz, & Schwarz, 1997; Snowden, Treue, Erickson, & Andersen, 1991). Thus the task of judging the direction of rotation of a transparent cylinder may not require SFM extraction. Instead, only separating motion signals in the transparent motion display may allow success in the task. Nonetheless, selective response to such complex motion displays is supportive of the putative role of this dorsal stream area in motion perception.

Selectivity to complex motion signals continues in up-stream dorsal areas. For example, MST cells show selectivity to oriented and tilted planes defined by SFM (Sugihara, Murakami, Shenoy, Andersen, & Komatsu, 2002). Cells in this region also exhibit selectivity for optic flow (Lappe, Bremmer, Pekel, Thiele, & Hoffmann, 1996), as well as object motion (Tanaka, Sugita, Moriya, & Saito, 1993). Remarkably, the selectivity for complex motion signals continues into the posterior parietal regions, and there is evidence for selectivity for SFM in these highest levels of dorsal stream of processing (Vanduffel et al., 2002).

A related body of evidence concerns the selectivity of posterior parietal regions for 3-D visual cues. Orban and colleagues (Durand et al., 2007; Orban et al., 2005; Orban, Janssen, & Vogels, 2006; Vanduffel et al., 2002) report in both monkeys and humans that posterior parietal cortex, particularly along the intraparietal sulcus (IPS) and even anterior portions of the IPS (AIP) are selective for 3-D cues including SFM and stereopsis. While some of the early data was obtained from FMRI studies, more recently single-unit recordings have corroborated the responsivity and selectivity of LIP neurons for complex shapes (Lehky & Sereno, 2007).

Effects of Lesion and Microstimulation of the Dorsal stream

Lesions of area V5/MT, while affecting pursuit eye movements in a retinotopically-specific manner (Dursteler & Wurtz, 1988), can also impair a monkey’s ability to detect coherent motion in a random-dot kinematogram (Newsome & Pare, 1988), when the coherent motion signal is placed in the retinotopic region affected by the lesion. Additionally, Newsome and colleagues (Murasugi, Salzman, &

Newsome, 1993; Salzman, Britten, & Newsome, 1990; Salzman, Murasugi, Britten, & Newsome, 1992) have shown that microstimulation of V5/MT can affect the judgment of motion direction. Lesions to V5/MT may impair 3-D SFM perception while leaving basic motion perception intact (Andersen & Siegel, 1990). Taken together, the results of temporary and permanent manipulations of V5/MT suggest a central role for this region in motion perception.

Lesions to the posterior parietal cortex also affect visual perception, namely spatial relations and 3-D perception as well. Holmes and Horax (1919) reported a patient suffering from bilateral posterior parietal damage due to a gunshot wound approximately in the region around the angular gyri. The patient exhibited normal object naming and reading, but was severely impaired at understanding spatial relations of objects or even remembering the paths he had taken. Pohl (1973) investigated effects of posterior parietal lesions in monkeys on visual perception and reported that such lesions impair the monkeys' ability to perform a landmark discrimination task that required comprehension of spatial proximity. Goodale and Milner (1992) have suggested that lesions of the posterior cortex in humans impair visually guided motor control, such as grasping and reaching. They proposed that optic ataxia, an impairment of visual guidance of actions brought about by damage to the posterior parietal cortex, and visual agnosia, impaired object recognition caused by lesion to the ventral stream, are suggestive of a double-dissociation. In this formulation, the posterior parietal cortex is essential for visual control of actions, such as grasping and manipulation.

Manipulation of posterior parietal cortex can also affect 3-D perception. The patient reported by Holmes and Horax (1919) complained of an inability to perceive depth in objects. For example, he would see a glass tumbler as “ a piece of flat glass” and would say about the man in front of him “I can only see the front of him, I do not notice that he is thick”. Tsutsui, Jiang, Yara, Sakata, and Taira (2001) report that temporary deactivation of the caudal intraparietal sulcus (CIP) by muscimol injections resulted in impaired discrimination of surface orientation in monkeys when the surfaces were disparity defined, and in one case, even if the surface orientation was defined by both disparity and linear perspective.

Summary

Taken together, the evidence from single-unit physiology, functional imaging, and lesion studies in humans and monkeys suggests the dorsal visual pathway plays an important role in the representation of visual motion, spatial perception, as well as use of 3-D cues such as disparity, perspective, and SFM.

Ventral Stream

Single-Unit Properties

Cells in V4, the first uniquely ventral stream region, are typically selective to colour and gross shape variables, such as length, width and curvature (Cheng, Hasegawa, Saleem, & Tanaka, 1994; Connor, Gallant, Preddie, & Van Essen, 1996; Desimone & Schein, 1987; Desimone, Schein, Moran, & Ungerleider, 1985; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; Pasupathy & Connor, 1999). Cells in the later stages of the ventral stream, areas TEO and TE on the inferior temporal gyrus, are far more selective in their response patterns (Desimone, Albright, Gross, & Bruce, 1984; Gross, Bender, & Rocha-Miranda, 1969; Perrett, Rolls, & Caan, 1982). The preferred stimuli of these regions are typically complex, and the complexity appears to be broken down and processed by individual columns (Fujita, Tanaka, Ito, & Cheng, 1992) suggesting a population coding of object features for a unified representation of the object. In such a scheme, the multiple features that make up an object, such as the ears and limbs on an animal figure, are each processed by columns selective for that visual feature, but the overall pattern of response amongst the columns represents a given object. This is supported by data from Tsunoda et al. (2001) using intrinsic signal optical imaging.

Another unique property of cells in IT is high selectivity for biologically salient objects such as faces (Bruce, Desimone, & Gross, 1981; Gross et al., 1969; Perrett et al., 1982). While initial reports of the proportion of IT cells that are face selective was rather low, these findings may have been influenced by various factors, such as the extensive use of anesthetized preparations. More recently, Tsao and colleagues (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Tsao, Freiwald, Tootell, & Livingstone, 2006) have suggested that the proportion of such cells may be as high as 97% if one spatially targets recording areas using FMRI-identified clusters. Tsao and colleagues used FMRI in awake fixating monkeys to identify patches of inferior temporal cortex that respond more to faces than to other objects. They then recorded from cells in the identified patches and found that the majority of cells in one patch observed in the middle of the inferior temporal gyrus was highly selective for faces, thus confirming the FMRI results. The researchers suggest this region to be a homolog of the human face-specific region termed the Fusiform Face area by (Kanwisher, McDermott, & Chun, 1997). However, an alternate interpretation may be that in the studies reported by Tsao and colleagues (Tsao et al., 2003; Tsao et al., 2006), when the monkeys viewed the same set of stimuli for several hundred hours, they formed new categories. Judging from the stimuli displayed, the non-face object categories used by Tsao et al. (2006) contained members that were highly dissimilar visually (i.e. an open hand and a fist together belonged in the “hand” category). Thus it could be said equally well that the “face” patches reported by those authors represent a “homogeneous category” effect rather than a “face category” effect.

It is not immediately clear how the columnarly-distributed pattern of representation as reported in IT cells relates to the finding that certain object categories such as faces may have a focal representation in the same regions. One hint comes from a study by Wang, Tanifuji, and Tanaka (1998) utilizing intrinsic signal optical imaging to identify the pattern of columns that represent multiple views of a 3-D schematic head. The researchers report a strip of IT tissue whose functional property can be broken down into a series of highly selective patches for different views of the head. This suggests at least two things—that patches responsive to an object category may be represented in adjacent columns and that multiple views of an object may be distributed across these adjacent columns. The former provides a bridge between focal category selectivity seen in IT while the latter is suggestive of a mechanism for view-invariant representation of objects in the brain. This latter issue is further discussed in Chapter 2.

Effects of Lesions and Microstimulation of the Ventral stream

Lesions to the ventral stream typically result in a variety of deficits that can be grossly categorized as deficits in form, object quality and object recognition. The reported "psychic blindness" by Klüver & Bucy (1937) was an early indication that the temporal lobe may play an important role in object recognition—following such large temporal lesions, monkeys appeared to misunderstand objects and attempt to eat every object placed in front of them. Mishkin (1954) selectively lesioned the temporal cortex and found that lesions in this region result in an object discrimination deficit. Further selective lesioning of different aspects of the ventral stream have largely corroborated the earlier findings. For example, lesions of V4 affect discrimination performance for complex patterns (De Weerd, Desimone, & Ungerleider, 1996; Heywood, Gadotti, & Cowey, 1992; Schiller, 1995), while similar observations have been also made after lesions of IT. Britten, Newsome, and Saunders (1992) found that while IT lesions do impair a monkey's ability to learn new 2-D forms from luminance cues alone, they do not prevent the animals from learning new forms defined by motion. This implies that while IT is important for form perception, it may not be necessary for all types of form perception.

Temporary deactivation of IT has supported an important role for this area in complex visual perception (Horel, Pytko-Joiner, Voytko, and Salsbury, 1986; Horel, 1996), while microstimulation of IT can serve as a paired cue in a visual associative task (Kawasaki and Sheinberg, 2008). Afraz, Kiani, and Esteky (2006) have evaluated the perceptual effect of IT microstimulation. They trained monkeys to perform a categorization task using different degrees of noise added to face images and found that microstimulation of face-selective patches biased the response of the monkey more towards the face category.

Summary

The evidence from single-unit physiology, intrinsic signal optical imaging, and FMRI along with lesion, temporary deactivation, and microstimulation of ventral visual areas in humans and monkeys together strongly suggest a central role for this region in complex visual perception and object recognition.

Correlated function and interaction of the two streams

While a number of functional dissociations may be suggested to exist between the two pathways as briefly reviewed above, there is much functional overlap as well. For example, although area MT/V5 may exhibit a high degree of motion direction selectivity, V4 cells may also exhibit direction selectivity if the motion is behaviouraly relevant (Ferrera, Rudolph, & Maunsell, 1994). V4 neurons that do not normally exhibit direction selectivity for motion may do so following adaptation to a motion stimulus (Tolias, Keliris, Smirnakis, & Logothetis, 2005). Cells in IT can be sensitive to higher-order disparity (Janssen, Vogels, & Orban, 2000) although the majority of disparity-selective neurons are found in the dorsal stream. These results suggest that many aspects of motion and spatial cue processing that were previously thought to be exclusive to the dorsal stream are also processed in the ventral stream.

Some qualities of ventral stream processes appear to be also present in the dorsal stream. Recently, Schlack and Albright (Schlack & Albright, 2007) reported that cells in MT can learn associations between static images and motion direction, developing shape selectivity for a static pattern that is associated with a direction of motion. This implies that V5/MT cells can also “learn” forms in a manner not totally dissimilar to the ventral stream processes. Kourtzi, Bulthoff, Erb, and Grodd (2002) have suggested that area MT can also discriminate between whole and scrambled objects, in the absence of motion cues.

Cells in both pathways may exhibit shape selectivity. Lehky and Sereno (2007), in the first direct comparison of shape selectivity in the dorsal and ventral stream, compared the response of TE cells and LIP cells to eight simple luminance defined patterns. Not only did LIP neurons respond well to these simple 2-D shapes, but also their responses were twice as strong and twice as fast as the TE neurons. Furthermore, the LIP cells, while as a population exhibiting lower selectivity for individual objects than TE cells nonetheless exhibited significant shape selectivity. While the authors speculated that their observations may have been different had they used 3-D shapes, it is interesting that even their simple 2-D forms that did not readily seem manipulable nonetheless activate a significant amount of dorsal stream neurons in a selective manner.

Extraction of surfaces from certain 3-D cues such as SFM and stereopsis appear to uniquely engage dorsal visual areas including MT, MST and posterior parietal regions such as LIP and CIP (Durand et al., 2007; Orban et al., 2005; Shikata et al., 2001; Shikata et al., 2003; Shikata, Tanaka, Nakamura, Taira, & Sakata, 1996; Sugihara et al., 2002). Indeed, while responses to shapes defined by 3-D stereopsis cues have been observed in ventral regions, such as cells in IT (Janssen et al., 2000), it is more likely that these regions receive related information from the dorsal stream regions that primarily computed the surface structure from a given cue. In other words, there is no evidence yet that ventral stream regions are directly involved in the computation of surface structures from 3-D depth cues.

The fact that dorsal visual areas are highly sensitive to 3-D cues and show shape selectivity for complex shapes is perplexing. One line of thought suggests that these dorsal stream 3-D representations are primarily (if not uniquely) for visually guided motor control, such as grasping and reaching (Goodale & Milner, 1992; Valyear, Culham, Sharif, Westwood, & Goodale, 2006). Another possibility, explored in Chapter 2, is that shape selectivity in the dorsal stream relates to normal object recognition. For example, dorsal stream lesions may impair 3-D SFM perception while leaving basic motion perception intact (Andersen & Siegel, 1990). The patient reported by Holmes and Horax (1919) had difficulties in perceiving the depth of objects due to damage to his posterior parietal cortex. Tsutsui et al (2001) reported impaired depth perception following temporary inactivation of caudal intraparietal sulcus. The present pattern of evidence thus suggests that dorsal stream regions may compute surface structures from 3-D depth cues such as SFM and stereopsis, and these extracted surfaces are then later related to ventral stream regions. In this way, dorsal input is critical for normal object recognition, and thus the mechanisms of object recognition can be considered distributed across the two pathways.

A number of questions are evoked by this discussion of dissociable and overlapping function of the two streams. First, can a complex task such as object recognition be constrained to one stream, as currently believed? Second, how does the shape selectivity of posterior parietal cells relate to their ventral stream analogs? Can object recognition take place exclusively in the dorsal stream? Finally, what is the relationship between the distribution of information across the two pathways and the current theories of object recognition and issues of complex perception such as view-invariance, familiarity and expertise?

Chapter 2

Dorsal-Ventral Integration in Object Recognition

The following paper attempts to synthesize findings in functional neuroimaging, animal electrophysiology, neuroanatomy, and psychophysics to better understand the nature and distribution of object recognition mechanisms in the brain. An important development has been the finding that posterior parietal neurons also exhibit shape and object selectivity and are central to the processing of many depth cues such as structure-from-motion and stereopsis. One view that explains these results is that the shape selectivity in the posterior parietal regions are important for planning actions, which may require 3-D understanding of objects (Goodale & Milner, 1992; James, Humphrey, Gati, Menon, & Goodale, 2002; Valyear et al., 2006). Alternatively, parietal shape selectivity may be related to object recognition. For example, dorsal stream regions, beginning with V5/MT, demonstrate selectivity to structure-from-motion (SFM), a monocular cue of depth that is derived from the differential displacement of surface details that are projected onto the retina when an object rotates in depth. So far it appears that extracting and computing surfaces from SFM is largely dependent on the dorsal stream. The fact that we can clearly see and recognize objects from SFM cues thus suggests that the two streams must interact in normal vision, because SFM cues are abundant in normal vision. This in turn places some constraints on the representation of object shape that can take place in the ventral stream. For example, ventral stream mechanisms must represent 3-D information, which can be considered a basic form of view-invariant representation. Additionally, the same mechanisms may likely be cue-invariant in order to allow for correspondence between multiple shape cues.

Dorsal-Ventral Integration in Object Recognition

Reza Farivar
McGill University, Montreal, Canada

Correspondence to:
Reza Farivar
Department of Psychology
McGill University
Montreal, QC, Canada
H3A 1B1
Reza.farivar@mail.mcgill.ca
Tel: (514) 398-6151
Fax: (514) 398-3255

Abstract

The idea of two parallel hierarchical pathways in vision has fueled a great deal of research and enhanced our understanding of visual processing in the brain. However, after 25 years, it has become clear that the earlier distinctions in terms of neuroanatomy and functional dissociation are less pure than originally considered. In the following review, I discuss research concerning the dorsal and ventral representations of object shape and attempt to integrate the results with models of object recognition. Based on current evidence, dorsal visual areas appear to play an important role in normal visual object recognition by computing surface structures from 3-D cues and providing this input to ventral visual regions for object recognition. This dorsal input to ventral areas is likely to be view-invariant and cue-invariant, empowering complex object recognition such as the recognition of faces from motion. It is proposed that normal object recognition is the result of the integrative action of the two streams.

Acknowledgements

I am grateful to Profs Ruffin Vogels and Guy Orban for their encouragement and comments on earlier drafts of this paper. This work was funded by a Doctoral Fellowship from the Fonds de la recherche en santé**.**

Introduction

In 1982, an idea was presented that dramatically influenced thinking about the primate visual system. Ungerlieder and Mishkin (Ungerleider & Mishkin, 1982), based on the pattern of behaviour following lesions to dorsal (occipito-parietal) and ventral (occipito-temporal) regions of the monkey cortex, suggested that the visual cortex can be decomposed into two pathways—a dorsal pathway concerned with spatial properties of vision (answering the question "where?") and the ventral pathway concerned with identification of the visual objects (answer the question "what?"). However, after 25 years, many challenges have been raised to that original elegant and simple view (Merigan & Maunsell, 1993; Hegde & Felleman, 2007), and an alternative description of the two pathways exists in terms of vision for perception (ventral stream) and vision for action (dorsal stream) (Goodale & Milner, 1992). While the original model and its variant still serve as useful paradigms for interpreting results from psychophysics, neurophysiology, neuroanatomy, neuropsychology, and functional imaging, they are still evolving to incorporate newer findings. The objective of this article is to highlight a number of studies that together suggest the two pathways are functionally integrated in normal object recognition to permit cue-invariant and viewpoint-invariant recognition. This may at first appear to contradict the original ideas of

Ungerlieder and Mishkin (1982) or those of Goodale and Milner (1992), but at closer inspection, it will be evident that normal object recognition and all the variable viewing conditions that may challenge it necessitate the integrative action of these two streams.

Models of Object Recognition

Models of visual object recognition can be divided along multiple, orthogonal dichotomies. The grandest dichotomy is between models that assume viewpoint-invariance in the neural representation of objects, and those that assume that viewpoint-invariant effects can be explained by uses of multiple individual viewpoints in an image-based manner. In the latter case, the brain interpolates intermediate views and thus allows us to recognize known objects from novel angles (Riesenhuber & Poggio, 2000). The viewpoint-invariant models suggest that the brain builds a structural representation of objects from available views, and this structural representation, analogous to a 3-D model, may be used to recognize the seen objects from novel views. While there is support for both models, the viewpoint-dependent models have the upper hand in explaining the vast majority of data obtained on representations of complex shapes, but in general, many agree that a combination of structural and image-based descriptions is necessary for normal object recognition (see Peissig and Tarr, 2007 for a review).

Category-level and subordinate-level recognition

An important aspect to consider in evaluating models of object recognition is to what extent they explain category-level and subordinate-level recognition performance. The human object recognition system must not only recognize objects as belonging to specific categories, such as “cat”, “chair”, “car”, et cetera, but must also be able to recognize individuals within that category—my neighbour’s cat, my car, etc. An area of object recognition research that informs us best about this aspect of object recognition is that of face recognition, a within-category type of object recognition that is essential to normal human engagement.

It is often argued that faces are processed differently than other objects, but it is unclear whether this difference is due to faces being processed by a specific “module” (Kanwisher et al., 1997), or by a general purpose visual object expertise system (Gauthier, Skudlarski, Gore, & Anderson, 2000). We are experts at recognizing faces, because this is a skill that is essential for our normal social interactions. Some of the effects observed uniquely for faces can also be observed for objects with which one has developed some expertise (Tarr & Cheng, 2003). For example, regions of the brain that respond more to faces than other objects also do so to objects that one has developed some expertise with (Gauthier et al., 2000; Xu, 2005). The electrophysiological correlate of face perception, N170, is significantly affected if another object of visual expertise is simultaneously presented (Rossion, Kung, & Tarr, 2004). While still a matter of considerable debate, there is evidence to suggest that aspects of face perception and recognition may be general to objects of visual expertise. How does visual expertise relate to models of object recognition?

Familiarity, expertise, and viewpoint-invariance

Booth & Rolls (1998) found that cells in the inferior temporal (IT) cortex of monkeys, a region that appears to be specialized for high-level object and face recognition (Gross et al., 1969), can demonstrate viewpoint-invariance for familiar objects, even when the developed familiarity is incidental rather than instructed. Monkeys in their experiments were given toys to play with in their home cages before the recordings. Of the 290 visually responsive cells that were found in the IT, 21 responded to these familiar objects in a viewpoint-invariant manner. Using information theoretic analysis, the authors show that the response of these 21 view-invariant cells together is sufficient to discriminate individual toys. Note that here, only visual familiarity was assessed, not visual expertise. However, one could presume that visual expertise would involve extensive familiarity with more than one member of an object class. If extensive familiarity with an object yields a more viewpoint-invariant representation, then one would expect that this level of viewpoint-invariance would influence the development of expertise with this category.

Whether such viewpoint-invariance holds for objects of visual expertise, such as faces, is less clear. Recognition of a face in a novel viewpoint is harder, both in accuracy and reaction time (Hill, Schyns, & Akamatsu, 1997; Troje & Kersten, 1999), although a recent study has suggested that the absence of 3-D information in tests of viewpoint-dependence may explain some of the detrimental effects of viewpoint change (Burke, Taubert, & Higman, 2007). Face adaptation effects appear to hold across large changes in viewing angles (Jiang, Blanz, & O'Toole, 2006). Face adaptation effects are an example of high-level visual after-effects (Leopold, O'Toole, Vetter, & Blanz, 2001) whereby following extensive viewing of an adaptor face, the perception of a mean face is biased towards an “anti-face” and vice-versa. This after-effect lends support to multidimensional face-space models, whereby each face is represented as a point in multidimensional space, with an average face describing the center of this space. If one imagines a vector originating from the average face to the target face, one could extrapolate an anti-face by moving to the opposite side of the mean on the same trajectory. The fact that face after-effects generalize across viewpoints suggests that face representations have high potential for viewpoint-invariance, although Jiang et al. (2006) suggest that this aftereffect may be generalizable to the case of other objects. Thus it may be the case that extensive experience with faces may have honed our ability to be able to guess what someone looks like from a novel angle.

Additional support for the relationship between view-invariant representation and familiarity comes from another high-level aftereffect—the viewpoint aftereffect. This phenomenon occurs when an object such as a face is viewed at a particular angle for a prolonged period of time. Following such adaptation, the perception of a frontal view appears skewed towards a face turned to the opposite direction (Fang & He, 2005). Ryu and Chaudhuri (2006) have shown that such aftereffects decrease with increased familiarity with a face. This suggests that following familiarity, a face is encoded in a less viewpoint-specific manner—or put in another way, familiar faces are represented in a more view-invariant manner.

Cue-invariant representation in the dorsal and ventral streams

Are viewpoint-invariant representations analogous to 3-D descriptions? One would expect a cell that is selective to different views of a given object may represent some aspect of the 3-D structure of the object that is projecting this image. What exactly this knowledge may be is an altogether different question, but whatever code is used to represent objects, it must be generalizable to situations where an object is not directly derived from image patterns, as in the case of objects defined by pure stereo, pure structure-from-motion, or a combination of the two depth cues. To reduce redundancy and increase efficiency, we would expect that viewpoint-invariant representations would be cue-invariant as well, and at least the results of Jiang et al. (2006) suggest this to be the case. The alternative is multiple representations of an object, each pertaining to one of the depth cues, working in concert to represent the visual input.

To what extent are object recognition mechanisms cue-invariant? There is evidence to suggest that cells in the inferior bank of the superior temporal sulcus on the inferior temporal gyrus are sensitive to higher-order disparity, such as the disparity one may perceive in natural viewing of objects. Janssen and colleagues (2000) suggest that this high-level disparity may have a dorsal origin because IT cortex receives input from regions in the IPS. However, another 3-D cue to depth structure, structure-from-motion, appears to be computed in “higher” regions in the dorsal stream hierarchy suggesting object recognition from this cue may represent dorsal-ventral integration.

Structure-from-motion cues are generated whenever objects on the retina rotate in depth, either from the movement of the head about the object or the rotation of the object in depth. The differential velocities of points on the surface enable the derivation of the three-dimensional structure, although the derivations are never exact (see Andersen and Bradley, 1998, for a review). Furthermore, our perception of 3-D SFM is similar to that of other primates, allowing us to investigate the neural basis of SFM in an animal model (Siegel & Andersen, 1988). While the exact locus of the computation of SFM is not clear (if there is even such a thing), a great deal of evidence suggests that SFM selectivity does not emerge before area MT (Grunewald et al., 2002), but may involve a number of other dorsal areas, such as MST (Sugihara et al., 2002) and posterior parietal areas (Vanduffel et al., 2002). It also appears that humans and monkeys do differ in their cortical representation of SFM although behaviourally they perform similarly (Orban et al., 2005; Vanduffel et al., 2002).

What is the evidence against earlier processing of SFM, such as at the level of V1? Grunewald, and colleagues (2002) assessed the neural response of V1 and MT cells in awake behaving monkeys that were trained to report their percept in a bi-stable SFM cylinder. This stimulus consists of a transparent cylinder that rotates about its long axis and its surfaces are defined solely by SFM. Because of the transparency in the cylinder, the stimulus is bi-stable—it could be equally perceived to rotate in one direction or the opposite, depending on the perceptual ordering of the two surfaces. In order to ensure that the monkeys are correctly reporting their percept, on some trials additional disparity information was available that guaranteed a specific depth ordering of the surfaces. These catch trials ensured that the subjects were reporting their true percept, thus enabling one to investigate the neurophysiological basis for the perception of 3-D SFM. Grunewald et al. (2002) found that whereas the response of about 20% of V1 cells is modulated by the reported percept, the responses of over 60% of MT cells are modulated with the reported percept. Furthermore, the modulation in V1 was not correlated with the cells direction tuning whereas the modulation in MT cells was correlated, suggesting that MT cells are carrying out the bulk of the computation of the SFM surface, and V1 modulation is a result of feedback from the MT cells.

Area MST, the satellite of MT, also shows particular selectivity to SFM stimuli. According to Sugihara and colleagues (2002), cells in MST have selectivity to the orientation, tilt and slant of SFM-defined planes. The anterior superior temporal polysensory area, a point of convergence between the dorsal and ventral streams, also appears to contain a number of cells that are SFM selective—they respond more strongly to dynamic motion patterns that describe a 3-D shape rotating in depth than controlled dynamic stimuli that do not elicit such a representation, but do contain motion nonetheless (Anderson & Siegel, 2005). In addition, human imaging studies suggest SFM is computed over a number of dorsal areas spanning the occipital, posterior temporal and the posterior parietal cortices (Andersen & Bradley, 1998; Orban et al., 2005; Peuskens et al., 2004). Importantly, ventral regions appear to be uninvolved in the derivation of surfaces from motion.

Dorsal stream regions exhibit selectivity to 3-D depth information as well as the 3-D spatial orientation of objects. For example, in addition to the aforementioned role of this pathway in the processing of SFM cues, dorsal regions such as LIP and CIP are also selective for stereo-defined patterns (Shikata et al., 1996; see Orban et al., 2006, for a review) and even anterior portions of the IPS complex may demonstrate 3-D shape selectivity in the monkey (Durand et al., 2007). Dorsal regions seem to be more selective to the orientation of objects in space and not their identity, while ventral regions are sensitive to their identity and less their orientation, as measured by repetition-suppression FMRI (Valyear et al., 2006). Thus it seems that dorsal stream mechanisms may be involved in aspects of object recognition such as for the perception of objects in space, or objects defined by depth cues such as SFM or stereopsis.

Dorsal-ventral integration in object recognition

While plenty of data now exists to suggest objects and shapes are indeed represented dorsally and certain 3-D cues of shapes are uniquely computed in dorsal-stream mechanisms, it seems clear that what we normally consider object recognition takes place in the ventral cortex. A number of issues then require clarification. First, how does the shape selectivity of neurons in the dorsal stream relate to object recognition in the ventral stream? Second, what is the nature of the object representation in the ventral stream that allows for integration of multiple cues about objects? Third, what is the interaction between familiarity, expertise, and the integration of multiple shape cues?

Shape selectivity in the dorsal stream

Lehky and Sereno (2007), in the first attempt at directly comparing shape selectivity in the dorsal and ventral stream, report some interesting results. In two awake fixating monkeys, they tested the selectivity of neurons in the lateral intraparietal (LIP) and IT cortex for eight simple black-on-white forms. They found that LIP neurons responded to the patterns faster than IT neurons (~60 ms versus ~100 ms), they responded almost twice as strongly, and had a more sustained response to the stimuli. Furthermore, the LIP neurons exhibited slightly greater adaptation effects. In contrast, IT neurons exhibited greater selectivity and less noise in representing the patterns. These results imply that a great deal of shape representation could take place in dorsal regions, but fine discrimination between shapes, the kind that is required for successful object recognition, likely utilizes the higher selectivity of ventral regions for object representation.

It is important to note that Lehky and Sereno (2007) did not evaluate the response of the cells to 3-D shapes, which may have given different results. Peuskens et al. (2004), in a FMRI study that used task manipulations instead of stimulus manipulations to examine the cortical representation of 3-D motion, 3-D shape, and 3-D texture, found that while attention to 3-D motion activated more dorsal regions and attention to 3-D texture engaged more ventral regions, attending to 3-D shape activated both significantly, suggesting an interplay of dorsal and ventral stream mechanisms. More recently, Durand and colleagues (2007), using FMRI in awake, fixating animals, found that area LIP, along with the anterior portion of the intraparietal cortex, respond both to 2-D and 3-D shapes. It remains to be known how the 3-D representations in these higher-level dorsal regions compare with representations in their ventral counterparts.

Viewpoint-dependence and invariance in the two streams

Dorsal and ventral regions may also differ in their viewpoint-dependence and invariance. Valyear et al. (2006) compared the selectivity of dorsal and ventral regions for object identity and object orientation using FMR adaptation. FMR adaptation is the reduction in the BOLD response that is observed when a stimulus is repeated. This adaptation is taken as an index of the selectivity of a voxel to the stimulus—if the voxel is insensitive to a particular dimension of a class of stimuli, it will show adaptation when that dimension is changed. Conversely, a lack of adaptation suggests sensitivity to changes on the manipulated dimension (Grill-Spector & Malach, 2001; but see Sawamura, Orban, & Vogels, 2006).

Valyear et al. (2006) utilized this property of the BOLD signal in an event-related design that consisted of trials where two masked stimuli were either different in identity, different in orientation, both, or neither. They found a sharp dissociation between dorsal and ventral regions in their response to orientation and identity differences. Whereas a dorsal region identified in their analysis, comprised of the superior temporal gyrus and the posterior intraparietal sulcus, was only sensitive to changes in orientation, they found a ventral region on the junction of the occipital and temporal cortices on the posterior fusiform gyrus to be sensitive only to changes in identity and insensitive to changes in orientation. More recently, Konen and Kastner (2008), have replicated and extended the results of Valyear et al. (2006) and also using BOLD adaptation, found that both dorsal and ventral regions exhibit selectivity for object shapes and size invariance. Interestingly, Konen and Kastner (2008) found two regions in the IPS that also exhibited viewpoint-invariance in addition to size invariance. These results suggest that while dorsal areas, representing 3-D information, are sensitive to the orientation of the object in space, ventral areas that may be the putative recipients of the dorsal input are not. Somewhere during the communication between the two pathways viewpoint-invariance is achieved.

Taken together, the studies that have directly assessed the response to shape in the dorsal and ventral stream seem to suggest that dorsal regions do encode certain aspects of the objects, but this representation may be limited to encoding either the extent of objects in space, as in the case of 3-D representations, or the orientation of objects in space. It is unclear if and how this extensive object representation in the dorsal stream affects object recognition in the ventral stream.

Dorsal input to ventral viewpoint-invariant representations

To better understand how these two systems may exchange information about objects, it is useful to return to the discussion of

SFM and stereopsis cues of object shape, as surfaces from these cues are largely derived by processes in the dorsal stream and possibly combined there as well (Welchman, Deubelius, Conrad, Bulthoff, & Kourtzi, 2005). As described above, SFM cues in particular are believed to be computed largely by dorsal stream mechanisms that manifest themselves at the level of MT and downstream regions. So far, there is no evidence to suggest that the surface cues from motion are computed in the ventral stream, yet we are apt at recognizing objects from SFM, and can even carry out unfamiliar face recognition from SFM cues alone (Farivar, Blanke, and Chaudhuri, submitted). Thus the ventral stream must be capable of understanding these cues of depth that give rise to 3-D representation.

Based on the studies on viewpoint-dependence and invariance of IT neurons and selectivity to 3-D shape and 3-D structural cues in the dorsal stream, I suggest that the dorsal stream mechanisms that compute 3-D object structures for purposes other than perception and recognition nonetheless relate those representations to the viewpoint-invariant mechanisms in the ventral stream. This implies that objects learned from pure 3-D cues such as SFM and stereopsis will inevitably be represented in a more viewpoint-invariant manner than those learned from 2-D images. Such a view would predict, for example, more view-invariant cells than view-dependent cells would be found in the IT cortex of monkeys that are trained to discriminate between objects defined by SFM or stereopsis alone. There is already some indirect evidence to suggest that some aspects of this view may be true. Availability of stereopsis reduces the latency and improves the accuracy in recognizing an unfamiliar face from a novel viewpoint (Burke et al., 2007). But the bulk of the argument in this paper rests on the fact that purely surface-structural descriptions, as mediated by dorsal visual areas that represent 3-D cues of depth, do empower complex object recognition and even unfamiliar face recognition.

A prediction that derives from the above discussion is that viewpoint-invariant representations will be cue-invariant—given that more than one dorsally-processed cue may contribute to the perception of 3-D shape, then the ventral 3-D representations must process the surface information whether the input is coming from stereopsis-computing circuits or from SFM-computing ones and thus be cue-invariant in their response.

Three-dimensional representations drive familiarity, improve recognition, and empower expertise

How does the dorsal input relate to processes of familiarity and visual object expertise? Jiang, Blanz, and O'Toole (2007) have already suggested, based on their finding that familiarity with faces reduces the viewpoint-dependent effects, that viewpoint-invariance may serve as a useful index of visual familiarity. Extending this view one would expect that viewpoint-invariant representations would be the ultimate goal of ventral stream mechanisms (Rolls, 2004), and in conjunction with our notion of dorsal-ventral integration, one may imagine that objects defined by pure stereo cues or SFM cues would result in a paradoxical “instant familiarity”. This need not be the case. Any given view of an object contains a wealth of additional information about the object in addition to its 3-D structure. These include the texture of the object, its colour, and various other random surface details that enable one to dissociate one object from another highly similar object. While the dorsal input from SFM or stereopsis computations can be informative of the 3-D shape, they have no way of informing about the texture, colour, or other properties of the surfaces. Full familiarity with a natural object may thus include both the 3-D information about the object, which can be provided from non-image based information, and descriptions of the surface properties of objects, which are image-based.

A related question is whether 3-D information aids object recognition. A number of studies on unfamiliar face recognition have addressed this issue with mixed results. While rigid head motion may improve recognition of unfamiliar faces slightly (Pike, Kemp, Towell, & Phillips, 1997), it is unclear whether this is due to the availability of 3-

D information from SFM or from the availability of multiple views (O'Toole, Roark, & Abdi, 2002). Also, while initially it was found that stereo viewing of unfamiliar faces does not improve their recognition, a more recent study suggests that stereo viewing does help when other manipulations may degrade performance, such as perspective changes (Liu & Ward, 2006). This latter notion, that 3-D information helps when other cues fail, may serve as a good heuristic in understanding the interplay between 3-D cues and 2-D views. While the latter is richer in detail that allows better discrimination between similar members of an object class, the former is less sensitive to spatial manipulations or variations in surface properties. Thus while I suggest that familiarity and dorsally computed surface descriptions both result in viewpoint-invariant representations, this does not imply that viewpoint-invariant representations alone imply familiarity.

Given that visual expertise with a class of objects requires extensive familiarity with many class members, it is important to consider how visual expertise measures within the present view described above. If we accept the hypothesis that familiarity results in enhanced viewpoint-invariance, then it might be expected to find an association between visual expertise, surface descriptors, and viewpoint-invariance. As described previously, it seems that identity aftereffects do translate to novel viewpoints, and by extension, one could assume that visual expertise with faces is behind this. Extensive experience with many members of the face class may have resulted in a very robust extrapolation mechanism that allows one to generalize identity across viewpoints (Jiang et al., 2006). In this scheme, the brain not only learns to reduce the dimensionality of the space defining all faces, but does so across viewpoints using 3-D information. Again the viewpoint-invariant representations formed during familiarity may aid in arriving at this highly compact and powerful mode of representation—instead of the face-space reduction process taking place over a large set of viewpoint-dependent representations, it may be carried out over the smaller set of viewpoint-invariant tokens. This implies that visual object expertise can only develop if a sufficient amount of viewpoint-invariant representation of members of an object class exists.

Neuropsychological and functional neuroimaging evidence

Such a view places a great deal of emphasis on 3-D representations for recognition, even for within-category discrimination problems such as face recognition. But this emphasis may be warranted. For example, face recognition may involve 3-D surface understanding to a greater extent than usually granted. Inverting the contrast polarity of faces can impair their recognition even when image-based edge information is preserved (Galper, 1970). Note that in contrast polarity reversal, the curvature of the face become difficult to understand, and understanding surface curvature may be essential for face recognition. This is supported by the finding that prosopagnosia without object recognition deficits also results in an inability to match two abstract “amoeba” 3-D shapes—curved surfaces in depth (Laeng & Caviness, 2001). Such a result places a great weight on the need for 3-D curvature of a face for its recognition.

There is also some neuroimaging and neuropsychological data to support this conjecture. As noted above, Valyear et al. (2006) have found a region ventral to the area LOC that appears to be insensitive to changes in orientation of objects in space, but is sensitive to their identity. This is also supported by a recent fmr-adaptation study by Weigelt, et al. (2007), where they find that this region shows adaptation to apparent motion of a two-frame object rotating in depth, even when the test stimulus is a novel view of the same object. The region identified by that study corresponds to the ROI region of LOC in the study of Welchman et al. (2005)—this region appears to be sensitive only to the perceived 3-D shape of the stimuli in a cue-invariant manner. We thus have two of the criteria for a visual expert system as outlined above in this region, namely viewpoint-invariance and cue-invariance. But is this region critical for within-category discrimination such as face recognition? Evidence from acquired prosopagnosia suggests this to be the case.

Patient P.S., initially reported by Rossion on colleagues (2003) is a rare case of a “pure” prosopagnosic. She is severely impaired at face recognition but retains excellent object recognition performance. FMRI analysis of her cortical response to faces has shown that she retains a normally functioning fusiform “face area” (Kanwisher, 1997), because her lesion is located in the occipito-temporal junction including portions of the lateral occipital and the posterior fusiform—regions that almost perfectly correspond to the viewpoint-invariant regions described by Valyear et al. (2006) and cue-invariant regions described by Welchmann et al. (2005). However, this is not an isolated case. Steeves et al. (2006) report a similar finding in patient D.F., studied extensively by Goodale and colleagues. While she maintains normal FFA activity in response to faces compared to objects, she is impaired at discriminating between previously studied faces and new faces, as well as in naming famous faces. Taken together, the neuroimaging results, combined with the effects of lesions, suggest that the ventral region of the lateral occipital complex is essential for face recognition and coincidentally, this region responds in a viewpoint- and cue-invariant manner. These results are consistent with the notion that visual expertise derives from extensive familiarity with objects, which in turn results in increased viewpoint-invariant representations that are also cue-invariant in encoding 3-D shapes.

What about the dorsal input from stereopsis and SFM? Gilaie-Dotan, et al. (2002) report that both LOC and the posterior portion of the fusiform respond to 3-D shapes defined by stereopsis. We have reported results from an imaging study that supports this conjecture. In our study subjects passively viewed 3-D SFM faces and 3-D SFM chairs, as well as scrambled versions of each that contained equal motion information, but no meaning. We report (Farivar, Germann, Petrides, Blanke, & Chaudhuri, 2006) that SFM faces engage the ventral portion of the LOC complex more than SFM chairs or scrambled versions of each, without selectively engaging the middle fusiform face region.

Conclusion

Taken together, it would appear that while the neuroanatomical dissociations do exist between a dorsal and ventral visual pathway, interpretations of the functions of these streams is less certain. Specific tasks such as object and face recognition may not be subserved exclusively by ventral stream mechanisms, and there is some emerging evidence to suggest that certain aspects of object recognition, such as recognition of an object’s orientation in space, may be processed by dorsal stream mechanisms (Priftis, Rusconi, Umilta, & Zorzi, 2003). Priftis and colleagues report a patient with posterior parietal lobe damage that seems to have a specific inability to discriminate between mirror stimuli and their normal orientation. In their study, the mirror stimuli were generated by rotating on both x, y, and z axes (the z axis being the axis formed from the subject's eyes to the screen). Oddly, this patient was only impaired on mirror images formed from rotation on the y axis. Furthermore, the patient had no difficulties in recognizing the object, nor did he have any difficulties in preparing grabbing gestures that were dependent on the perceived orientation of the stimulus. He simply could not perceive which stimuli were mirror inverted on the y axis. It is noteworthy that most other studies of orientation change have used rotation about the y axis to probe orientation differences, and these were found to be more selectively represented in the posterior parietal cortex than in the ventral regions (Valyear et al., 2006). The fact that only one such patient study reports a visuo-perceptual deficit following dorsal stream damage highlights a bias in the field. As argued above, dorsal stream mechanisms may be more integral to visual perception than previously thought, and may be directly implicated in object recognition mechanisms that are thought to be purely "ventral". However, very few attempts have been made to identify perceptual dysfunctions following dorsal stream lesions. The fact that the first direct comparison of shape selectivity in the LIP and IT was conducted only this past year (Lehky & Sereno, 2007) speaks to the inherent bias in our view of dissociated function of the two streams.

The case for dorsal-ventral integration in vision is arguably clearer in the case of particular 3-D shape cues, such as stereopsis and SFM in particular, that appear to be largely computed in dorsal stream mechanisms. Structure-from-motion appears to be largely computed in high-level dorsal stream areas, namely MT and upstream, with currently no known aspect of the computation—the extraction of a surface from motion—taking place in the ventral stream. The fact that we can recognize and learn objects from 3-D SFM cues implies that dorsal-ventral integration takes place for this task. This in turn implies that the representation of objects in the brain must be flexible to include 3-D coding of object shape from various cues, resulting in quasi view-invariant and cue-invariant representations. The nature of such 3-D representations, their relationship to perception, memory formation, familiarity and expertise are important links to consider in our attempt to understand the neural mechanisms of object recognition.

Chapter 3

DORSAL-VENTRAL INTEGRATION IN THE RECOGNITION OF MOTION-DEFINED UNFAMILIAR FACES

Structure-from-motion can serve as a powerful cue to 3-D shape, but it is unclear whether this cue can empower complex object recognition such as the recognition of unfamiliar faces—arguably the most difficult object recognition task. Additionally, a model of cortical mechanisms of face recognition by O'Toole and colleagues (2002) suggests that SFM cues ought to be able to serve as an input to the ventral stream face recognition mechanisms. In contrast, findings from Britten et al., (1992) suggest that shape discrimination from dynamically generated forms may not be a ventral stream process. In this paper we present evidence that 3-D SFM cues can empower complex object recognition as measured by the recognition of unfamiliar faces defined by SFM cues. While we confirm that dorsal stream mechanisms are necessary for successful extraction of surfaces from cues, ventral stream areas appear essential for the recognition faces defined by SFM, suggesting that dorsal-ventral integration is necessary for the recognition of SFM-defined unfamiliar faces.

Dorsal-Ventral Integration in the Recognition of Motion-Defined Unfamiliar Faces

Abbreviated Title: Recognizing Motion-Defined Unfamiliar Faces

Reza Farivar[1*], Olaf Blanke[2], & Avi Chaudhuri[1]

1 Department of Psychology
McGill University
1205 Doctor Penfield Ave.
Montreal, QC, H3A 1B1

2 Laboratory of Cognitive Neuroscience (LNCO)
Brain-Mind Institute
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne

Keywords: unfamiliar face recognition, dorsal-ventral integration

Acknowledgements

This work was supported by FRSQ and NSERC fellowships to RF, CIHR operating grants to AC, and grants from the Swiss National Science Foundation and Fondation de la Famille Sandoz to OB. Special thanks to Eugene Mayer, Michela Adriani and Karen Borrmann for their time and help.

Abstract

The primate visual system is organized into two parallel anatomical pathways, both originating in early visual areas but terminating in posterior parietal or inferior temporal regions. Classically, these two pathways have been thought to subserve spatial vision and visual guided actions (dorsal pathway) and object identification (ventral pathway). However, evidence is accumulating that dorsal visual areas may also represent many aspects of object shape in absence of demands for attention or action. Dorsal visual areas exhibit selectivity for 3-D cues of depth and are considered necessary for the extraction of surfaces from depth cues and can carry out mnemonic functions with such cues as well. These results suggest that dorsal visual areas may participate in object recognition, but it is unclear to what capacity. Here we tested whether 3-D structure-from-motion (SFM) cues, thought to be computed exclusively by dorsal stream mechanisms, are sufficient to drive complex object recognition. We then tested whether recognition of such stimuli relies on dorsal stream mechanisms alone, or whether dorsal-ventral integration is invoked. Results from a prosopagnosic patient confirm that ventral stream areas are necessary for both identification and learning of unfamiliar faces from SFM cues, while such cues are sufficient to drive unfamiliar face recognition in normals.

Introduction

The cortical visual areas of primates are broadly organized into two separate anatomical pathways, a dorsal pathway that includes areas in the posterior parietal cortex (PPC) and a ventral pathway that includes inferior temporal (IT) regions (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). The two pathways have been thought to represent different aspects of vision—the dorsal pathway representing spatial relations and visually guided actions and the ventral pathway being critical for object identification.

While ventral visual areas are considered important for complex visual object recognition, many aspects of object recognition may also be carried out and replicated by visual areas in the PPC. Lehky and Sereno (2007) found that cells in areas LIP of the monkey responded strongly and rapidly to 2-D forms with a pattern similar to IT cells recorded in the same study. Konen and Kastner (2008) using fmr-adaptation in humans, report two areas along the intraparietal sulcus (IPS) that showed adaptation to 2-D forms and 3-D shapes, regardless of the object's viewpoint or size. Size- and viewpoint-invariance are essential for an object recognition system and dorsal visual areas exhibit these properties (James et al., 2002; Valyear et al., 2006).

Visual areas in the PPC in humans and monkeys exhibit selectivity for 3-D cues of shape such as structure-from-motion (SFM), stereopsis, and perspective (Shikata et al., 1996; Shikata et al., 2001; Sugihara et al., 2002; Shikata et al., 2003; Anderson and Siegel, 2005; Orban et al., 2005; Durand et al., 2007). Some of the 3-D cue-selective neurons in these regions exhibit properties that are suggestive of a role in "high-level" visual perception. Cells in the caudal IPS (CIP) exhibit orientation-selective and delay-sustained activity during delayed matching of two 3-D oriented surfaces (Tsutsui et al., 2003). Furthermore, temporary deactivation of this area results in impairment on this discrimination task (Tsutsui et al., 2001). These results imply that dorsal visual areas are involved in certain cognitive aspects of shape processing from 3-D cues.

The processing of visual motion is commonly thought to depend on dorsal stream mechanisms as well. Dynamic aspects of a visual scene provide important cues for object segregation and identification. For example, gestures, emotional expressions, and idiosyncratic head movements can be used to drive identity and gender categorization in the absence of other shape cues (Hill and Johnston, 2001). On the other hand, 3-D SFM cues can be derived from all visual objects. These cues are highly informative of object shape and may be capable of driving complex recognition processes in the absence of other shape cues or idiosyncratic movements.

A number of attempts have been made to estimate the contribution of SFM to face recognition (O'Toole et al., 2002). However previous studies had not separated the sole contribution of object motion from monocular cues (e.g., shading) or other motion cues (e.g., facial gestures and identity signatures). Although a specific role for SFM has been postulated by a model of face recognition (O'Toole et al., 2002), to date no direct evidence exists in support of this model.

Experimental Procedures

We first sought to assess whether naïve observers can utilize SFM cues to carry out a complex object recognition task—namely recognize unfamiliar faces. We then attempted to distinguish between the two competing hypotheses outlined above—one postulating a role for dorsal visual areas in object recognition from 3-D cues, and the other postulating the necessity of dorsal visual areas for the extraction of surfaces from depth and the ventral visual areas for the recognition and identification of the 3-D objects.

Our stimuli consisted of 3-D laser-scanned heads (Troje and Bulthoff, 1996) and 3-D models of chairs and other objects that were rendered using a unique texture mapping technique (3-D procedural texture mapping). This approach eliminates sources of biological motion as well as monocular depth cues such as shading and texture gradients. The resulting images have no defining 2-D features that may be used to recognize the objects (Figure 1). The motion-defined objects are invisible when the display is static. However, rotating the surfaces in depth yields a vivid 3-D percept from the SFM cues.

Stimuli and Design

Three-dimensional laser-scanned heads from the Max Planck database were used for these experiments (Troje and Bulthoff, 1996). The stimuli were rendered with 3-D procedural texture maps to ensure uniform textures, as described in detail in our previous work (Liu et al., 2005). The 20 heads rotated in depth from left to right, from -22.5º to 22.5º about the vertical axis at a rate of 27.3º/sec, and were rendered with perspective transformation. The recognition targets were the same heads, but rendered with shading only, in orthographic projection to avoid simple metric matching. Twenty subjects participated in each of the first two experiments (mean age 26.8, 15 females and 25 males), with ten in each condition. Subjects viewed the rotating SFM faces that extended approx. 30º of visual angle vertically and 21º horizontally and identified the face amongst eight gender-matched targets. All participants gave written informed consent before inclusion in the study, which had been approved by the Research Ethics Board of McGill University (Canada) and the Ethical Committee of the University Hospital of Geneva (Switzerland).

Patient Studies

Information on the patients is provided in Table 1. First, all patients viewed a series of 15 objects and 3-D geometric shapes defined by SFM and were asked to name them. Following this, all patients completed a series of additional 1:8 identification tasks as described above, consisting of rotating SFM faces and rotating SFM chairs (rendered in a manner identical to the faces). Finally, their ability to match static displays was tested on the same task but with static shaded faces and chairs.

Patient P.S., suffering from prosopagnosia, completed two additional tasks designed to probe her capacity to use SFM cues for face and object discrimination and recognition. She completed a face-learning task where she was required to learn to name four faces (two male, two female) presented via SFM. Each of the faces was present for 3.3 seconds only (one rotation), and she was encouraged to respond as fast as possible. The patient viewed each of the faces 80 times in the course of the study. Her residual ability at object recognition was also tested using a chair-learning task that was carried out in the same manner as the face-learning task.

Controls

We measured the performance on the 1:8 identification tasks on both patients with lesions that left their vision unaffected and normal subjects with no neurological damage. Two control patients, one suffering from damage to temporal and parietal cortices and exhibiting aphasia and the other suffering from damage to the parietal cortex participated in the same tasks described above. In addition, eight subjects (5 females and 3 males), aged 46-52 (mean=50.1, SD=2.1) with no neurological impairments and normal or corrected-to-normal vision participated in the matching tasks. Eight normal age-matched normal controls underwent the additional face and chair learning tasks that P.S. completed, with four in each object category condition.

Results

We first sought to learn whether naïve subjects are able to recognize unfamiliar faces defined by SFM. A previous study had suggested that SFM cues may be of limited use in *familiar* face recognition, but are not sufficient for *unfamiliar* face recognition (Bruce and Valentine, 1988). It remains unclear whether facial movement in general (as in the case of continuous multi-view video of a face) aids better recognition than a single photograph (Pike et al., 1997; Christie and Bruce, 1998). This type of rigid movement would include SFM cues along with other cues, thus it would not speak directly to a role for SFM in face recognition. Although at least one model of cortical object processing suggests a role for SFM cues in face recognition (O'Toole et al., 2002) , there is no direct evidence to validate this claim.

In the first experiment, subjects viewed motion-defined face stimuli on one screen while attempting to identify the face amongst eight choices (target faces) on another screen. The eight target faces were rendered as static shaded faces, similar to sculptures, and were matched for gender with the motion-defined face. One group of subjects viewed the dynamic faces, whereas another group viewed a single static frame. This condition served as a control to ensure that there were no contaminating factors in the stimuli that could aid face recognition in the absence of dynamic information. We found that subjects viewing the control stimuli performed at chance (Figure 2) whereas subjects viewing the SFM faces performed approximately four times above chance ($t(18) = 5.9916$, $p < 0.0001$).

We next tested whether transient texture gradients formed while the face rotates in depth can be used for successful recognition. The same recognition task was used, but with textures that rotated incongruently with head rotation. These stimuli could therefore only be recognized if the transient texture gradients served as a reliable source of structural information, given that SFM cues were removed. Subjects in this condition performed slightly above chance (Figure 2) but significantly below the SFM group ($t(18) = 4.5097$, $p < 0.001$). Together, these results confirm the usefulness of purely dynamic cues of shape, devoid of other monocular depth cues or biological motion signals, in driving complex object recognition such as the recognition of unfamiliar faces.

We next sought to distinguish between the two hypotheses outlined above, concerning the role of the dorsal 3-D representations in object recognition. If 3-D shape representations in dorsal visual areas were sufficient to carry out complex visual object recognition, then a patient with ventral stream impairment would have no difficulty on tasks requiring identification and object learning from 3-D cues such as SFM. If, on the other hand, dorsal 3-D shape representations must be relayed to ventral stream regions for object recognition, as postulated by O'Toole et al. (2002), then ventral-stream impairment would be the limiting factor for successful recognition of shapes from 3-D cues such as SFM. We tested these contrasting possibilities in neuropsychological cases of akinetopsia (Zihl et al., 1983) and prosopagnosia (Damasio et al., 1982). The former represents an impairment of dorsal stream visual processing resulting in impaired motion perception whereas the latter represents impairment in the ventral stream to produce a specific inability to recognize faces.

Patient V.D. is a 47-year old, right-handed man suffering from akinetopsia due to Alzheimer's disease (Rizzo and Nawrot, 1998). He exhibited a severe impairment for direction discrimination from coherent motion and orientation discrimination of 2-D forms-from-motion. However, he did not have any object recognition deficits, and so we were interested to know if he could use 3-D SFM cues by a system other than his impaired dorsal stream. Additionally, we were interested to know if there was any other information in our 3-D SFM stimuli beside the motion-defined structure that could be used to drive discrimination performance, even though earlier control studies had suggested this to not be the case. In effect, the performance of Patient V.D. served as a negative control for the stimuli and paradigm used here.

The results from this patient, shown in Figure 3, suggest that he is unable to extract motion cues from the displays and thus unable to perceive motion-defined stimuli. It is unlikely that non-motion cues were present in the stimuli because otherwise he would use this information to drive his performance above chance. However, he can recognize stimuli if they are defined by other cues, such as shading, suggesting that he does not have a difficulty making fine discriminations. The fact that his near normal performance with the shaded stimuli did not translate to any residual ability to perceive the 3-D SFM stimuli confirms that the extraction of surfaces from these dynamic cues requires putatively dorsal stream mechanisms.

To assess the necessity of ventral stream structures in the recognition of motion-defined stimuli, we examined Patient P.S., whose clinical condition was previously studied in detail and reported by Rossion et al. (2003). P.S. is a 57-year old right-handed woman who suffers from prosopagnosia. She exhibited no difficulty in perceiving SFM stimuli and performed perfectly on the object-naming task. On the 1:8 identification tasks, her performance replicated some of the earlier reports using face and object photographs by Rossion et al. (2003). Her identification accuracy with the chairs, while not as good as normal controls, was well above chance and within 2 standard deviations of the normal performance (Figure 3). However, she was impaired on face identification—her 1:8 matching performance with SFM faces was at chance and more than 2 standard deviations below the group average. With shaded faces she was able to perform above chance, but still significantly worse than the normal controls. She has developed a strategy of using the lips to match faces, and this facial feature is difficult to identify in the right-to-left rotating SFM faces, but clear in the shaded stimuli. Thus it is likely that her strategy of using the lips drove her performance on the shaded faces above chance, but her performance was still more than 2 standard deviations below the normal control group.

We additionally designed a task to test her capacity to learn unfamiliar motion-defined faces and motion-defined chairs. She was asked to learn the names of four faces (two male and two female), four office chairs, or four armchairs that were such that the set of armchairs were similar in homogeneity to the set of faces. P.S. was unable to learn the faces even after 80 repetitions of each face, whereas four age-matched controls were able to reliably learn the task (Figure 4a). Her raw performance for each face across the sessions is displayed in Figure 4b. In contrast to normal controls, her performance is unreliable over time—the occurrence of correct and incorrect responses for each face is random. She reported facility at perceiving the face and all of the facial components but, similar to face photographs, she reported that she could not “put the face together”. Performance on a similar chair-naming task (Figure 5a) remained unaffected. Her performance with motion-defined office chairs reached a ceiling after only ten trials and was comparable to her performance with the shaded stimuli. When we used highly similar chairs (armchairs), her performance increased more slowly, but she was clearly able to learn the chairs as evidenced by her consecutively correct performance on the chairs and the similarity between her performance and that of age-matched normal controls (Figure 5b). Thus the chair- and face-naming tasks were similar in task difficulty as evidenced by the similarity in the performance of normal controls on the two tasks.

Discussion

We have shown that ventral stream mechanisms are necessary for complex object recognition using SFM cues even though the bulk of evidence suggests that dorsal stream mechanisms are essential for extracting the surface structure from this depth cue. Our results lead to several conjectures.

First, the results from the naïve subjects suggest that motion cues alone are sufficient to drive complex object recognition including the recognition of unfamiliar faces. This may at first stand at odds with studies that suggest head motion does not enhance face recognition, but note that here only 3-D SFM cues were available, not additional edge and shading cues. Thus it may be the case that SFM cues may not improve face recognition if other reliable cues are present. Liu and Ward (2006) found that a 3-D cue such as stereopsis improved face recognition performance when perspective transformation degraded performance. Thus it maybe the case that head motion may also improve recognition, but if 3-D perception is affected by a spatial transformation.

Second, the data from patient P.S. suggest that the ventral stream object representations are cue-invariant—that they may process a given object regardless of the 3-D cue used to define the shape. This is supported by the finding that the P.S. displayed a specific impairment that was category-selective for faces, but not cue-selective—she performed significantly worse than normal controls on both matching tasks with faces defined by SFM and those defined by shading. Her results imply that the ventral face processing mechanisms that she lacks were also the recipient of a putative dorsal input.

Third, although there is evidence that neurons in the PPC (e.g., area CIP) may represent 3-D surface information during delay periods (Tsutsui et al., 2001; Tsutsui et al., 2003), these mnemonic functions are insufficient for creating new memory associations for long-term reference. This is supported by P.S.'s inability to learn name associations to four faces from SFM. P.S. does not have a long-term or short-term memory deficit (Rossion, et al. 2003), thus her inability to learn the four faces from SFM-based stimuli is likely due to her category-selective impairment.

O'Toole et al. (2002) postulated a role for dorsal-ventral integration from SFM cues, though no direct evidence for this link had been provided until now. Kriegeskorte et al. (2003) found support for the O'Toole et al. (2002) model in an event-related paradigm with a face detection task that used two SFM-defined faces. While they reported increased FFA activity in response to faces compared to random surfaces, they found a similar category selective response even in the human homolog of MT (hMT+) as well as a differential response in IPS for faces defined by another type of motion cue (termed on-surface SFM). While their results suggest a role for the FFA in perception of motion-defined faces, the same role can be equally attributed to the hMT+ and IPS peaks observed in their study. Recently, Konen and Kastner (2008) have demonstrated, using an fmr-adaptation paradigm, that PPC shape selectivity is comparable to that of the ventral stream, thus highlighting the need to clarify the role of the dorsal stream shape representations in object recognition.

There is a growing body of evidence to suggest that dynamic cues such as SFM are processed by dorsal stream areas (Andersen and Bradley, 1998; Anderson and Siegel, 2005; Orban et al., 2005) whereas recognition of complex objects, such as faces, is dependent on ventral-stream processing (Haxby et al., 1991; Kanwisher et al., 1997; Ishai et al., 1999). Interestingly, monkeys with lesions to a specific part of the ventral stream—the inferotemporal cortex (area IT)—are unable to perform perceptual and memory-related tasks with luminance-defined patterns, but perform normally on perceptual tasks utilizing motion-defined patterns (Britten et al., 1992). Thus it appears that not all aspects of complex visual recognition depend on ventral stream mechanisms.

Ventral stream areas, such as the inferior temporal (IT) cortex of monkeys, are highly interconnected with parahippocampal areas (Seltzer and Pandya, 1991), leading to the conjecture that this cortical stream is important for memory formation and object recognition. Neural processes underlying perception of motion-defined patterns presumably remain undisturbed following ventral stream dysfunction. In humans, a ventral system impairment (agnosia) does not impair the ability to use motion-parallax cues for depth reach planning in a delayed-response task that requires retention of perceptual information (Dijkerman et al., 1999). Although dorsal stream areas may exhibit shape selectivity (Shikata et al., 1996; Nakamura et al., 2001; Lehky and Sereno, 2007), our results suggest that these regions may not be involved in object recognition *per se*, in the sense of allowing for comparisons to stored representations.

Our results have both neurobiological and clinical significance. It remains unclear whether dorsal-ventral integration requires synchronized activity between the two streams (Singer, 1999) and what exactly is the nature of the representation that is transmitted from dorsal stream areas to their ventral stream counterparts. The SFM-defined face recognition task also provides a novel probe of dorsal-ventral integration, allowing for studies on the role of attention in cortical integration or its disruption in neurological disorders.

Figure Captions

Figure 1. Generating purely motion-defined faces

(**A)** We used 3-D laser scanned heads to isolate structure-from-motion information and remove other cues, such as biological motion, shading, and texture cues. The stimuli were devoid of unique identifiers such as blemishes or distinct skin textures. (**B**) Using volumetric texture mapping, we generated a uniform density random dot surface on the head, analogous to carving the head out of a block of stone. A schematically low-density texture is used here to facilitate description. (**C**) Shading was eliminated by setting the object to illuminate like a lamp, and removing reflectance cues from the object texture. (**D**) Object boundaries were made invisible by placing the object in front of an equally high-density textured plane. The co-occurrence of the target on the textured background made it impossible to dissociate the object using only 2-D boundary information. Panel (**E**) depicts the final stimulus using high-density textures. No facial information is available in any single frame, but between any two frames the displacement of the texture in depth yields a vivid sensation of structure-from-motion.

Figure 2. Recognition performance of naïve subjects with unfamiliar faces defined by SFM and control stimuli.

In each condition, 10 participants performed the 1:8 identification task with SFM-defined faces (SFM), rotating heads with incongruent surface dot motion (Incongruent), and static frames of one of the SFM videos (Static). Participants who viewed SFM-defined faces performed well above chance, while participants who viewed the incongruent surface motion stimuli performed far worse than the SFM-viewing condition, but slightly better than the group that viewed the static frames of the SFM videos, who performed at chance levels.

Figure 3. Performance of the akinetopsic patient and prosopagnosic patient on a series of motion-defined and static, shaded stimuli

An akinetopsic (patient V.D.) and a prosopagnosic (patient P.S.) patient were tested on an identification task using structure-from-motion and structure-from-shading stimuli. In the naming task, the subjects were required to name a set of 15 objects and geometric shapes that were solely defined by motion. All other tasks were 1:8 identification tasks. SFM faces were more difficult in general to discriminate than chairs, though Patient V.D. was unable to perceive any of the SFM-defined objects. His performance on the matching task with shaded faces and chairs showed that form perception was not similarly affected. Patient P.S. displayed a specific inability to recognize faces. She had no difficulty naming SFM objects and matching SFM chairs, but her performance on SFM faces was at chance.

Figure 4. Performance of the prosopagnosic patient on face learning tasks

(A) Performance of Patient P.S. on a learning task consisting of only four motion-defined faces. Patient P.S. was unable to learn the faces reliably, though age- and gender-matched controls were able to learn the faces. **(B)** Although at times P.S. appears to perform above chance, her raw performance suggests otherwise. This panel depicts the raw performance of P.S. and normal controls on trials 21-60. Each column represents the response to a particular face while each row represents the trial number. Black cells represent an incorrect response; correct responses are shown by white cells. Performance across trials is inconsistent and she rarely identifies the same face correctly on consecutive trials, suggesting that she is not in fact learning the faces. Data from control subjects, however, suggest they all learned to name the faces.

Figure 5. Performance of the prosopagnosic patient on chair learning tasks

(a) In contrast to her learning performance with faces, P.S. had no difficulty learning motion-defined office chairs, suggesting that she does not suffer from a general impairment in recognition and learning of SFM-defined stimuli. **(b)** Here her raw performance on chair learning is re-assessed using four armchairs that were more similar, making the task equal in difficulty to the face naming task as evidenced by the normal control performance on the two task. Each column represents the response to a particular armchair while each row represents the trial number (trials 21-60). Black cells represent an incorrect response; correct responses are shown by white cells. Again, P.S. is able to learn the armchairs in a manner similar to the normal controls, in contrast to her face learning performance.

Table 1. Clinical details of the patients

Figure 1

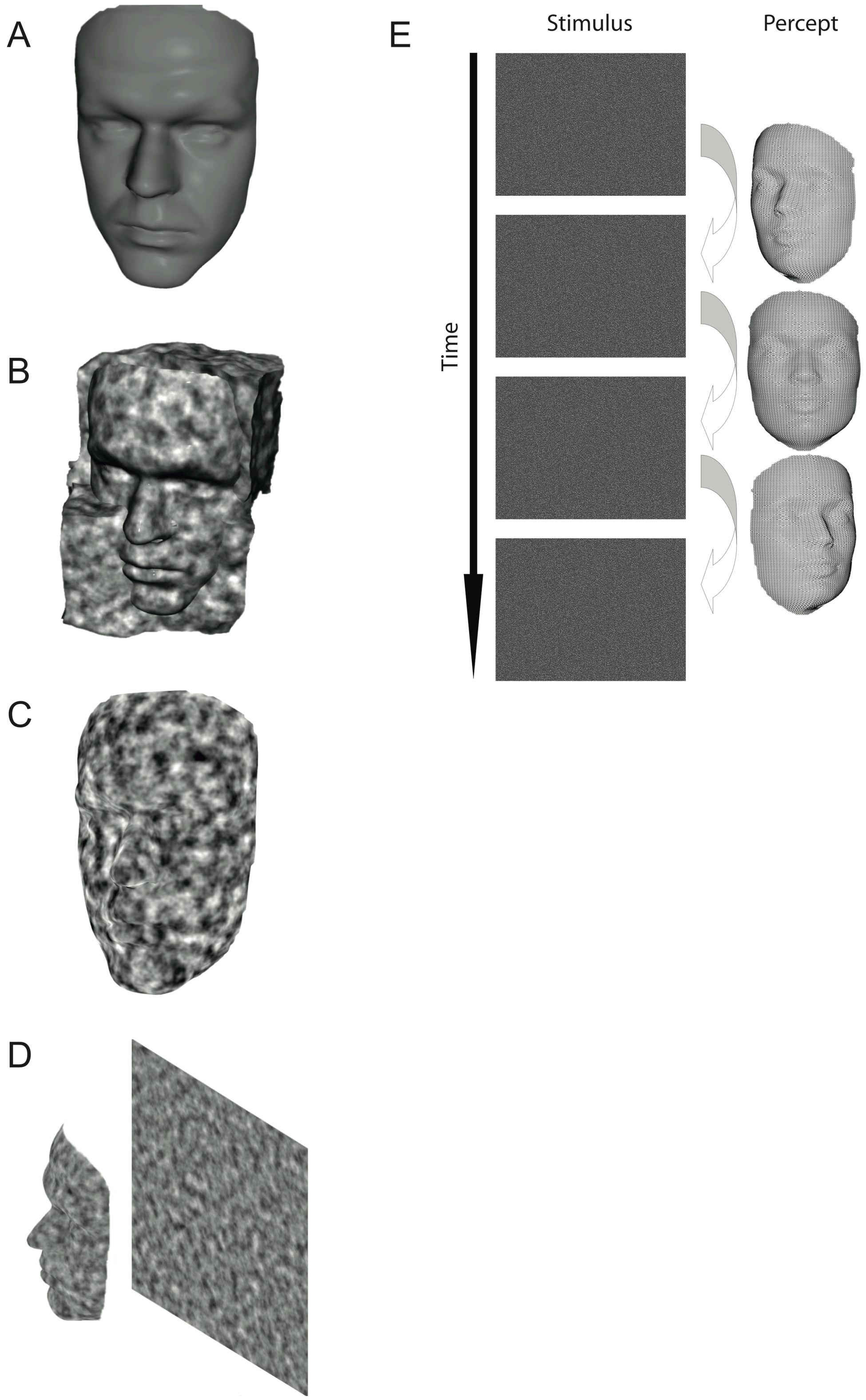


Figure 2

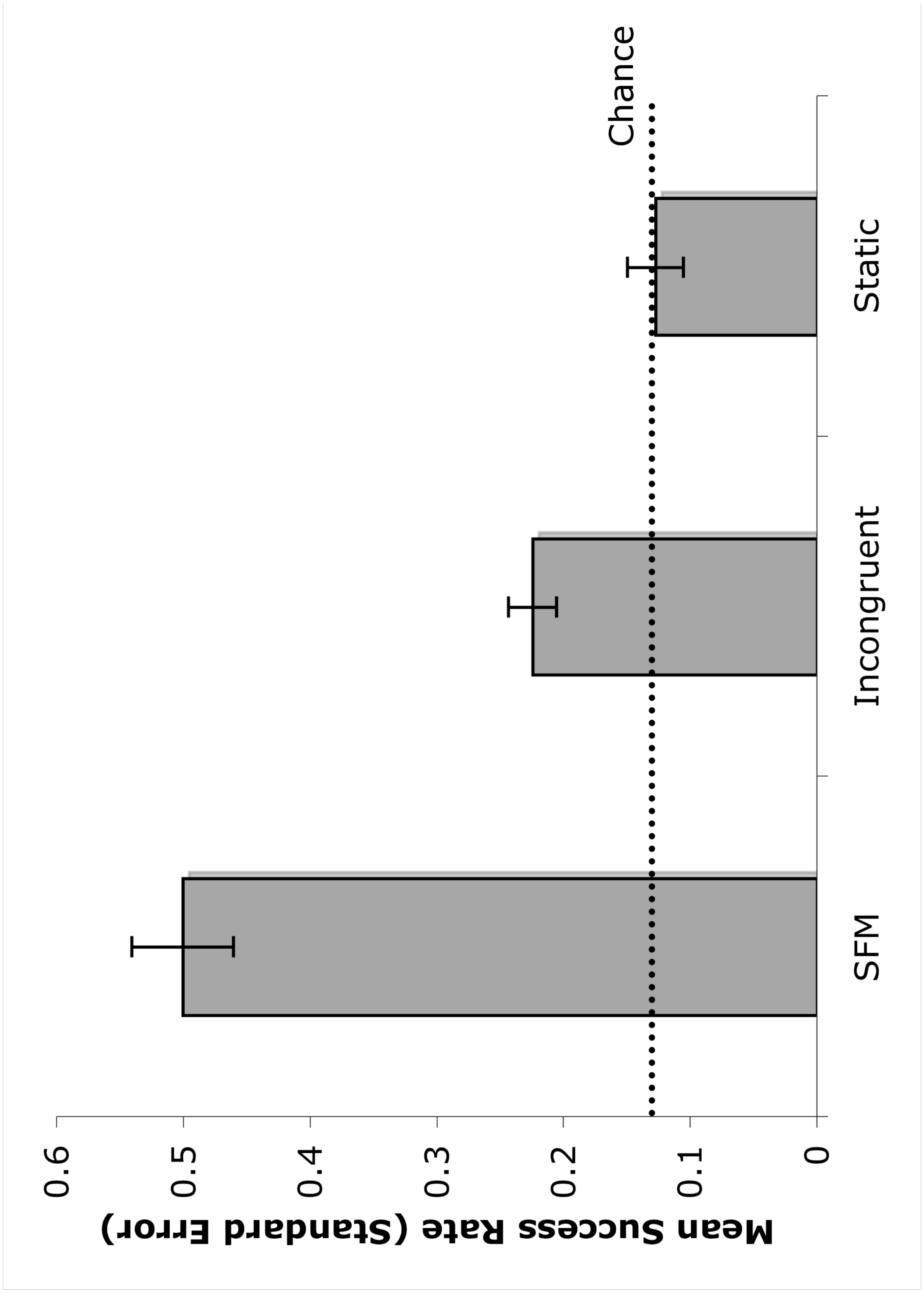

Mean Success Rate (Standard Error)
0.6
0.5
0.4
0.3
0.2
0.1
0
SFM
Incongruent
Static
Chance

Figure 3

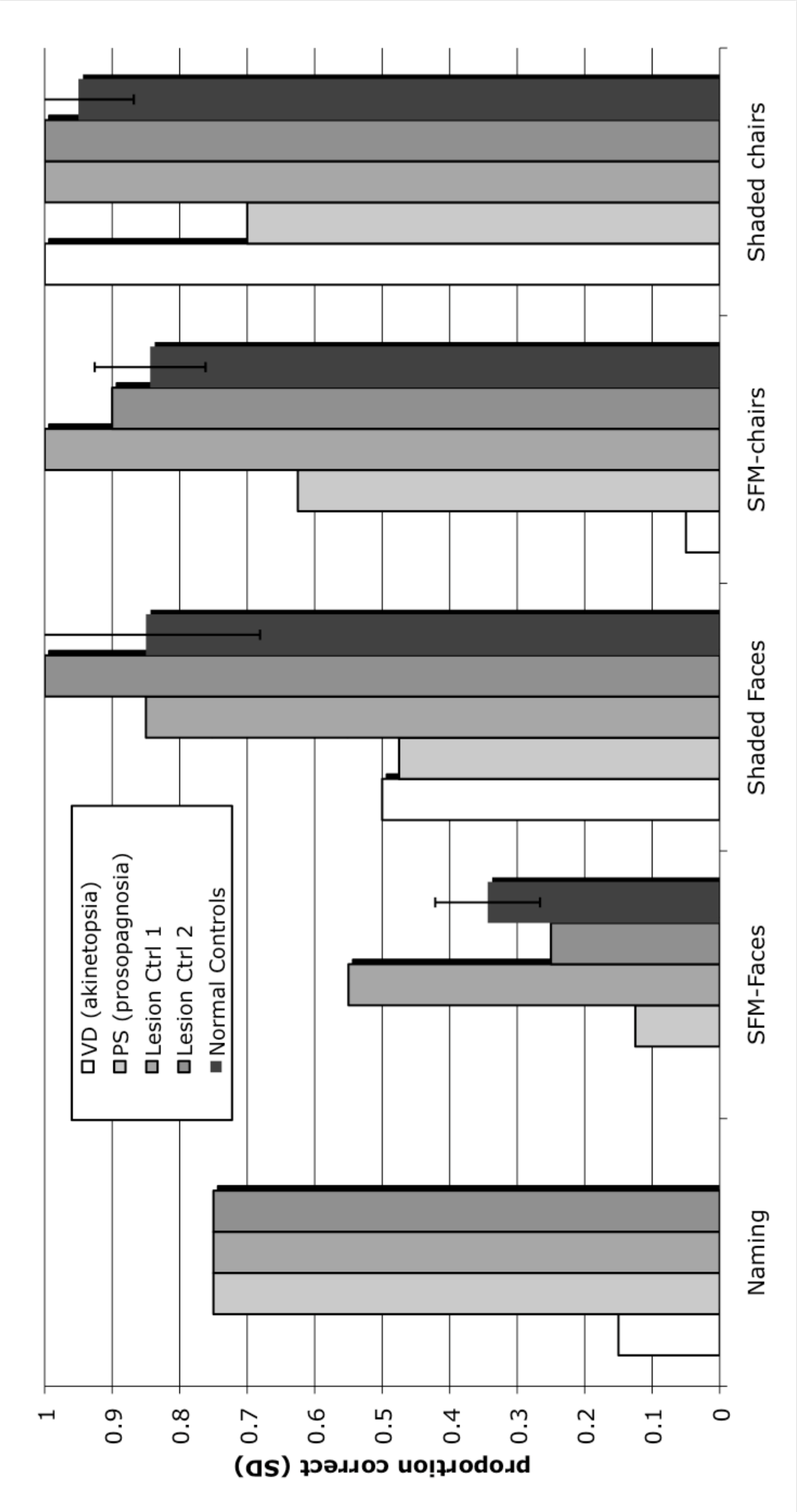

proportion correct (SD)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Naming
SFM-Faces
Shaded Faces
SFM-chairs
Shaded chairs
□VD (akinetopsia)
□PS (prosopagnosia)
□Lesion Ctrl 1
□Lesion Ctrl 2
■Normal Controls

Figure 4

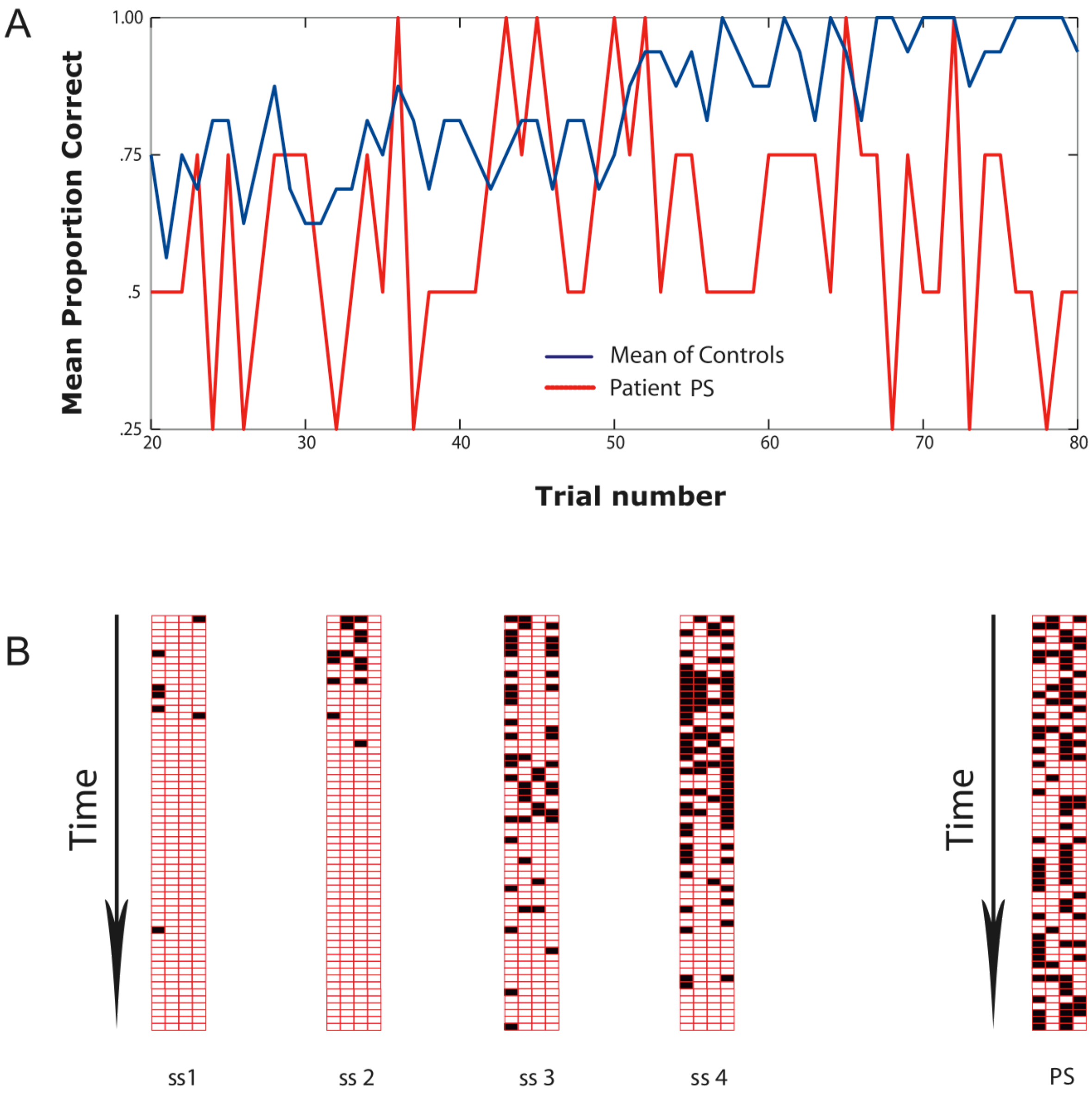

A
Mean Proportion Correct
1.00
.75
.5
.25
20
30
40
50
60
70
80
Trial number
Mean of Controls
Patient PS
B
Time
ss1
ss 2
ss 3
ss 4
Time
PS

Figure 5

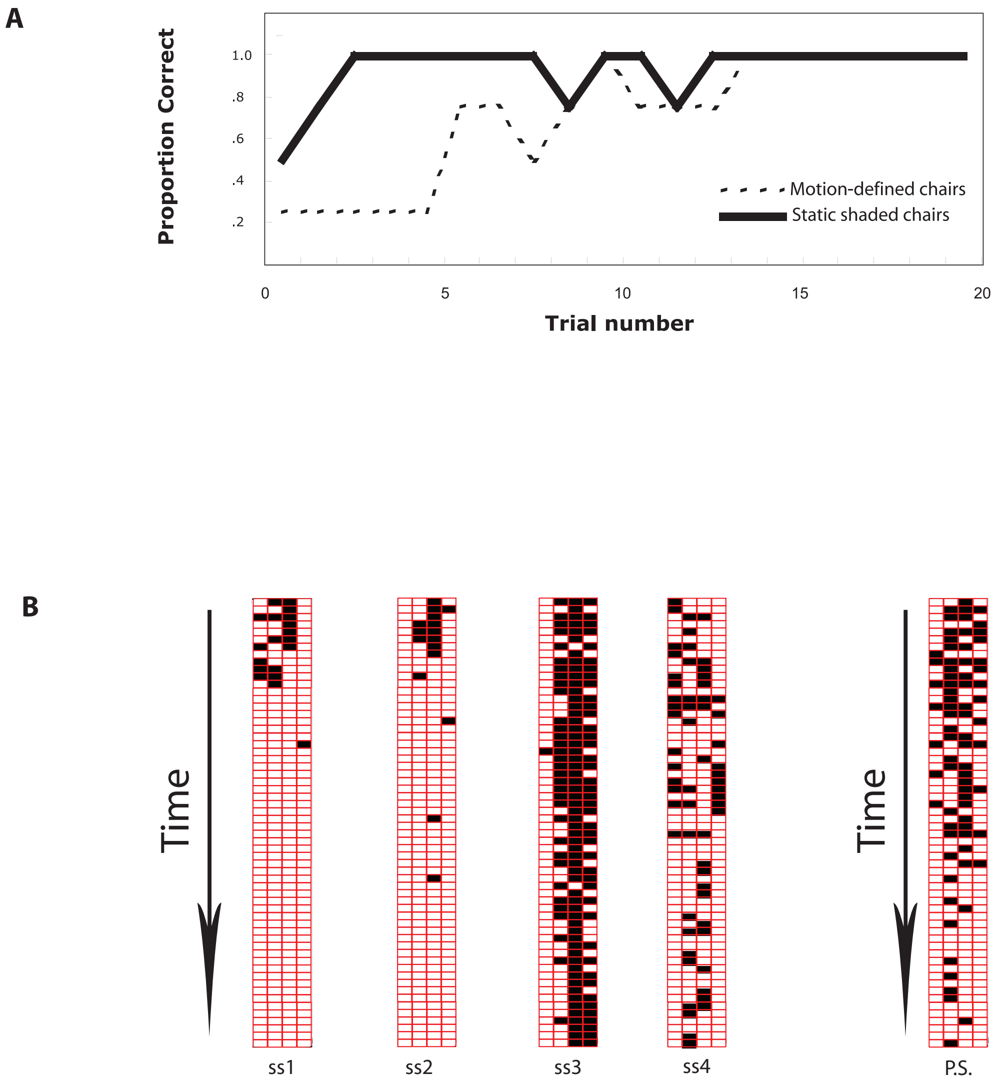

A
Proportion Correct
1.0
.8
.6
.4
.2
0
5
10
15
20
Trial number
Motion-defined chairs
Static shaded chairs
B
Time
ss1
ss2
ss3
ss4
Time
P.S.

Table 1

Patient	Age	Hem[1]	Lobe[2]	Etiology	Visual fields	Motion direction[3]	Form-from-Motion[3]
V.D.	47	B	TP	dementia	full	severe	severe
P.S.	57	B	TO	posttraumatic	full	normal	normal
Lesion Ctrl 1	21	R	P+	malformation, epilepsy	full	normal	normal
Lesion Ctrl 2	63	L	TP	stroke	full	normal	normal

[1] B—bilateral, R—right, L—left
[2] TP-Temporal & Parietal, TO-Temporal & Occipital, P+--Parietal plus white matter
[3] Severe—significantly elevated thresholds, typically at ceiling

(prosopagnosia)

Chapter 4

2-D Form-from-Motion Deficit with Intact 3-D Structure-from-Motion Perception

It is often suggested that the anatomical hierarchy of visual organization maps onto a functional hierarchy as well, with a distinction between low-level and high-level perception—the former encompassing simple detection and discrimination and the latter representing complex recognition, identification and categorization. In the following paper, we evaluate whether such a distinction is valid for simple and complex motion processing as exemplified by 2-D form-from-motion (FFM) and 3-D structure-from-motion perception.

Running Head: 2-D FFM versus 3-D SFM

2-D Form-from-Motion Deficit with Intact 3-D Structure-from-Motion Perception

Reza Farivar[1], Avi Chaudhuri[1], and Olaf Blanke[2]

[1] Department of Psychology, McGill University, Montreal, Canada
[2] Mind-Brain Institute, Ecole Polytechnique Federal de Lausanne, Switzerland

Correspondence to:

Reza Farivar
Dept. of Psychology
McGill University
1205 Doctor Penfield Ave.
Montreal, QC, H3A 1B1
reza.farivar@mail.mcgill.ca

Introduction

Besides telling us where things are going, motion can aid object segregation and recognition through 2-D form-from-motion (FFM) from kinetic grouping based on speed or direction or from 3-D structure-from-motion (SFM), a percept that results from the differential velocities of points on a surface rotating in depth. Whereas 2-D FFM can aid object segregation and recognition of objects from silhouette-type patterns, 3-D SFM is an important depth cue, present whenever viewers and/or objects move about in depth.

Vaina (1989) reported a dissociation of 2-D FFM from 3-D SFM processing deficits depending on lesion locations. Patients with lesions in the right occipito-temporal cortex showed an impairment of 2-D FFM and stereopsis but were normal in the processing of 3-D SFM, while patients with lesions to the right occipito-parietal were impaired in stereopsis and 3-D SFM but were normal in 2-D FFM perception. In a single case-study, Vaina et al. (1990) reported a patient who showed impairment in perception of coherent motion from random-dot kinematograms, speed discrimination and seeing 2-D FFM defined by relative speed of dots. Interestingly, this patient, suffering damage to the lateral parietal-temporal-occipital cortex, did not display a deficit in 3-D SFM. This is surprising because based on the current hierarchical view of object perception that begins with metrical comparisons in earlier areas and more complex 3-D perception later on, one would expect a correlated deficit in 2-D FFM and 3-D SFM, yet the results of Vaina et al. (1990) suggest that 2-D FFM and 3D-SFM are independent processes.

3-D SFM cues are abundant in our natural viewing. Each time our head moves with respect to an object, the relative displacement of the elements projected from the surface of the object on our retinas can provide cues to their shape, and this 3-D SFM cue can empower complex object recognition, including the recognition of unfamiliar faces defined by motion (Farivar et al., 2006). The recent findings on the capacity of 3-D SFM to enable complex object recognition was made possible by the use of recent 3-D computer rendering methods that allowed for the creation of complex object stimuli represented by SFM without any other 2-D cues such as shading or shadows. One difficulty in interpreting previous studies on SFM perception is that the task employed may not have required 3-D surface extraction. For example, 3-D SFM perception in studies by Vaina and colleagues (Vaina, 1989; Vaina et al., 1990) was assessed by means of a cylinder orientation task—a 3-D cylinder rotating in depth with variable number of dots on the surface and variable speed. Three points concerning this task merit consideration.

First, the projected rotating cylinder was transparent. The movement of transparent surfaces creates an additional challenge for the visual system, namely that it requires segregation of two different planes moving in opposite directions. Additionally, the stimulus is bi-stable and as such gives rise to competition for depth ordering between two surfaces moving in opposite direction. Moreover, the modifications made to the task in order to increase its difficulty (varying the number of dots or the lifetime of the dots) did not result in a degradation of performance in the 3-D SFM task. Accordingly, the 2-D FFM task used in Vaina (1989) may not be diagnostic of motion impairment, but could be explained by other factors, such as a form perception deficit (see Vaina et al., 1990, for a discussion). Also, detecting the direction of rotation of the transparent surfaces may allow one to successfully perform the cylinder rotation task without actual 3-D surface extraction. Taken together, it is plausible that the task as presented may not have been sensitive enough to tap into SFM deficits if they had existed, and where they are suggested to exist, the deficit could have been due to difficulties in segregating motion signals locally than estimating SFM per se.

Neuroimaging results using FMRI suggest that 2-D FFM and 3-D SFM processing may have different cortical substrates. Orban and colleagues (Dupont et al., 1997; Van Oostende, Sunaert, Van Hecke, Marchal, & Orban, 1997) report a region in the posterior occipital cortex, distinct from MT/V5 and V3 that they termed the Kinetic Occipital region (KO). KO appears to respond to kinetic boundaries and perhaps more specifically to figure-ground ordering by motion cues (Tyler, Likova, Kontsevich, & Wade, 2006), thus suggesting that it plays an important role in 2-D FFM perception. In contrast, Orban et al. (1999) and others (Murray, Olshausen, & Woods, 2003; Paradis et al., 2000) report a number of regions particularly along the intraparietal sulcus (IPS) that are selective to 3-D SFM cues. Thus it appears that human cortical regions relating to 3-D SFM processing are distinct from those related to 2-D FFM perception.

Here, we were interested in revisiting this issue in neurological patients with posterior brain damage, using novel SFM stimuli and tasks that require object recognition from 3D-SFM cues.

Below we describe a set of experiments conducted on two groups of patients selected on the basis of their functional impairment in motion perception as measured on a 2-D FFM task. Patients were also tested on a series of structure-from-motion and structure-from-shading identification tasks that probed the patients’ ability to perceive 3-D shapes defined by static cues and by dynamic cues. Our stimuli were designed such that their successful discrimination required accurate perception of surface curvature and relative depth. For example, one set of stimuli consisted solely of SFM-defined faces that could only be identified if the features of the faces could be discriminated accurately.

Methods

Participants

Patients

Four patients (two female), with impaired 2-D FFM perception (Blanke et al., 2007) participated in the study. Two additional patients (Patients 5 and 6) were control patients and also suffered from posterior cortical damage, but had no impairments on motion perception tasks. Further clinical details are provided in Table 1. Coherent motion thresholds and form-from-motion thresholds are shown in Table 2.

Healthy Controls

Eight aged-matched healthy control subjects (5 females; aged 46-52 years (mean=50.1, STD=2.1) with no neurological impairments and normal or corrected-to-normal vision participated in the study. Informed consent was obtained from all participants.

3-D SFM Tasks

Stimuli

For the object-naming task, fifteen 3-D objects that could be readily named were used (e.g. camera, umbrella, cube, etc.). For the face-matching tasks, 20 facial surfaces (10 male, 10 female) taken from the Max Planck database (Troje & Bulthoff, 1996) were used. Stimuli for the chair-matching tasks consisted of 20 chairs taken from commercial and free 3-D model databases. The same stimuli were used for both the structure-from-shading and structure-from-motion tasks.

The 3-D SFM stimuli were rendered in 3D Studio Max R3 (Autodesk) using our previous methods (Liu et al., 2005). Briefly, we used a random texture map with the following parameters (Threshold high=0.6, threshold low = 0.59). Each 640 x 480 pixel frame contained approximately 12000 dots ranging in size from 1 to 10 pixels. About 40% of the dots fell on the 3-D object surfaces (i.e., 4800 dots). Texture was applied procedurally, meaning that each texture element was generated individually on the object surface and was created to tile seamlessly with adjacent texture elements. Shading and shadows were not used. Objects were set to rotate in depth from left to right, from -22.5° to 22.5° about the vertical axis at a rate of 27.3°/sec. The resulting individual frames resembled random dot patterns and without motion, no object could be seen in the static frames.

3-D shaded stimuli and match targets were rendered with default 3D Studio Max shader (Blinn) with ambient light. For each object, one image was rendered by setting the orientation of the object 22.5° to the right in the same layout used for the SFM stimuli.

Apparatus

A dual-display (19”, 1280 x 1024 resolution, 60Hz refresh rate) PC computer running Matlab (Mathworks Inc.) was used to run the experiments. The subject viewed the target stimuli on a monitor placed in front of them (the primary display), at a distance of approx. 40 cm, yielding 52° of visual angle horizontally and 40° visual angle vertically. The eight match targets appeared on the other display in two rows of four, with an Arabic numeral identifier.

Procedure

Object Naming

As a basic diagnostic, subjects were asked to name 15 objects defined solely by 3-D SFM. These objects were readily namable, such as an umbrella, camera, airplane, as well as simple geometric shapes such as a pyramids, cones, etc. Subjects viewed these stimuli on the primary display and the experimenter recorded their verbal responses.

Identification Tasks

For all the 1:8 identification tasks the same procedures were used. Subjects viewed the target stimuli on the primary display and identified the match amongst the eight sample stimuli displayed on the secondary display. The task was not timed, but subjects were encouraged to respond accurately and promptly. In total, four 1:8 identification tasks were completed by the patients—(1) SFM face identification, (2) SFM chair identification, (3) shaded face identification, and (4) shaded chair identification.

2-D Motion Tasks

Coherent Motion Task

Coherent motion stimuli (random dot cinematograms, RDC) were presented on a 20″ computer monitor (Sony; frame rate, 70 Hz; 640 × 480 pixel resolution) in black and white in a normally lit room as described previously (Blanke et al., 2007). Viewing distance was 100 cm. The stimuli were presented in a borderless square of 12° × 12° in the central visual field. The percentage of coherent motion (%CM) was defined as the number of signal dots divided by the total number of dots and multiplied by 100. The remaining dots were noise dots and were plotted at random locations for a random duration (between 67 and 800 ms) giving the impression of flickering dots. Dots moving out of the stimulus area reappeared on the opposite side such that density was held constant. The direction of each RDC stimulus in each block was varied randomly between the four cardinal directions (right–left–up–down). An automated staircase algorithm varied the %CM in the RDC, starting at 100%CM (all dots moving in one direction).

Form-From-Motion Task

The 2-D FFM task was the same as the coherent motion task with respect to equipment, viewing distance, field of presentation, total number of dots, and dot velocity, as previously described (Blanke et al., 2007). A borderless, static form (a capital letter E) was plotted in the center of the 12° × 12° random dot field. The size of the letter was 6° × 6° and was defined by 250 signal dots with 750 dots outside of this central area. Percent coherence was calculated based on the 250 signal dots and for different coherence levels, different proportions of these 250 signal dots were converted to noise dots inside the central area of 6° × 6° for a random duration (between 67 and 800 ms) giving the impression of flickering dots as all dots outside the central area. Overall stimulus density was equivalent with the coherent motion task, thus ensuring task comparability. As in the coherent motion task, the direction of each RDC stimulus in each block was varied randomly between the four cardinal directions of motion and four orientations (right–left–up–down). An automated staircase algorithm varied the %CM in the RDC, starting at 100%CM (all dots moving in one direction).

Procedure

In the coherent motion task, subjects were asked to indicate verbally the direction of motion they perceived. In the FFM task, subjects were asked to indicate verbally the orientation of the letter E embedded in the stimulus. Feedback was not provided. Subjects provided their answer verbally and the examiner recorded the response. The rate of trial presentation was controlled by the examiner and adjusted to patient comfort. Subjects were instructed to look at the center of the screen and to refrain from making eye movements.

For each trial, the %CM increased or decreased depending on the performance of the subject on the last trial. Four independent staircases (one for each direction of motion) were randomly interleaved for both tasks. The four staircases were continued until five response reversals had occurred for each tested direction. The staircase steps had a scaling factor of 0.67. Thus, the %CM values (staircase steps) were generated using steps that decreased by 1/3 of the previous higher value (100 × 2/3 = 66.7; 66.7 × 2/3 = 44.4; 44.4 × 2/3 = 29.6; 29.6 × 2/3 = 19.8, etc.). The procedure (5 reversals) estimated about 0.8%CM for the coherent motion task and 8.7%CM in the FFM task in normal controls (Blanke et al., 2007). The mean of the last three reversals was taken as the %CM threshold. The mean of these four directional thresholds for each subject provided the threshold for coherent motion and FFM perception.

Results

Patients 1 through 4 were categorized as 3-D SFM impaired if they performed more than 2 standard deviations (SD) below the normal controls on the SFM-identification tasks. In this manner, Patients 1 and 2 were categorized as SFM impaired, while Patients 3 and 4 were not. Coincidentally, Patients 1 and 2 performed below chance (1/8, or 12.5%) on the identification tasks with SFM-defined objects, suggesting they could not use SFM information to make accurate identification.

Patient 1 performed nearly perfectly on the chair identification from shading cues, suggesting he can use 3-D information in static form to make the judgments. Despite his poorer performance with static shaded faces, his performance level was within 2 SD of normal performance for identifying shaded faces. In contrast, his SFM-defined object identification was below chance level, and his object naming performance was far worse than the control patients.

The performance of Patient 2 was generally worse than normal controls on all tasks, but was well above chance for identifying shaded static objects. Her performance with SFM-defined objects, however, was below chance, similar to Patient 1. Her object naming performance with SFM stimuli was also poor compared to control patients. Thus while she generally performed worse on all the tasks compared to normal, she appeared able to use shading to drive her performance to well-above chance levels, but not to normal levels of performance.

In contrast to Patients 1 and 2, Patients 3 and 4 performed the identification tasks with SFM-defined objects at the same level as the normal controls. Surprisingly, Patient 3 performed significantly worse than normal controls in identifying shaded static faces but well above chance. Her performance on shaded chairs, similar to her performance with SFM-defined chairs, was perfect, and her SFM-object naming was good as well, committing few errors.

Patient 4, on the other hand, performed normally on all identification tasks, within 2 SD of normal performance, and with perfect object naming performance. The lesion control patients all performed similarly to the normal age-match controls, suggesting neurological impairment *per se* was not responsible for the pattern of performance seen in Patients 1 through 4.

Although we selected Patients 1 through 4 based on their pattern of 2-D FFM impairment, we found that performance on the 2-D FFM task did not inform about the patients' performance on the 3-D identification tasks. However, Patients 1 and 2 also had elevated thresholds for detection of motion direction from coherent motion patterns (Table 2). This suggests that while 2-D FFM impairment does not correlate with 3-D SFM impairment, impairment of detection direction in coherent motion may be related to 3-D SFM impairment.

Discussion

Several points must be made clear concerning the 3-D shape tasks. First, the face recognition tasks were rather difficult. Recognition of faces devoid of 2-D facial features such as eyes, eyebrows, etc., is generally considerably harder than the recognition of faces from face photographs. Our previous results with shaded-only, texture-only, and SFM-only face recognition tasks suggest that matching performance on these is never 100% (Liu et al., 2005; Farivar et al., 2006). The performance of Patients 3 and 4 on the SFM-face naming task is comparable to that of the control Patients 5 and 6 and normal controls, whereas Patients 1 and 2 perform considerably worse (at zero), where chance is 1/8 (approximately 3/20 on each matching task). Given that all patients have severely elevated 2-D FFM thresholds, these data suggest that 3-D SFM performance is uncorrelated with 2-D FFM performance, but correlated with the performance in coherent motion perception.

One possible explanation for the poor performance of the patients 1 and 2 on the SFM task is that they could extract SFM-defined shapes but could not recognize the stimuli. Yet, this explanation is unlikely for several reasons. Both patients performed well above chance when the stimuli were presented in static shaded form. Even the SFM object-naming task that all other patients found quite easy were still challenging for these two patients. They also performed below chance on the SFM chair identification task, which other subjects found considerably easier. Also, the matching tasks did not make a heavy memory demand nor did they require naming the object, thus language or memory related deficits could not explain the poorer performance. Taken together, these data suggest that the severe 3-D SFM deficit of Patients 1 and 2 was not due to form perception deficits.

The approach we made in measuring 3-D SFM perception made use of the fact that many complex objects such as every day objects, faces, and chairs can be represented by 3-D SFM. Although we did not systematically vary a single variable to create a graded scale of performance, our tasks did broadly fall into easy, medium, and difficult tasks. For example, the patients (with the exception of the two SFM-impaired patients) found the object-naming task quite easy and reported the name of the object readily and without hesitation. On the chair naming tasks, these same patients were able to match the chairs to their shaded targets expediently but with some consideration, given the fact that chairs have many elements in common. Yet these same patients expended considerably more energy on the face matching tasks, probably due to the high degree of similarity between different faces in general and faces of one gender in particular. In this manner we were able to roughly gauge the level of SFM performance in these patients.

Our study employed a functional criterion of inclusion rather than a neurological criterion based on lesion location. As such, we cannot speak directly to the cortical mechanisms responsible for 2-D FFM and 3-D SFM. However, given that patients 1 and 2 both suffer from temporo-parietal damage, the results suggest that parietal regions may play a greater role in the perception of 3-D SFM than temporal regions. The key difference between the 2-D FFM and 3-D tasks may be the nature of local computations necessary. Although both sets of tasks require integration of motion across space and grouping of elements based on dynamic features such as direction, 3-D SFM recognition requires comparison of velocities across space for the estimation of the rotating surface.

The key finding that we report here is a lack of relationship between 2-D FFM performance and 3-D SFM performance, but an association with impaired coherent motion perception. This is surprising, because both the 2-D FFM and 3-D SFM tasks require grouping of similar elements (motion in a given direction) and integration across space for the correct generation of the percept.

Furthermore, one would assume, based on a hierarchical model of object recognition, that 2-D form perception would precede 3-D object recognition. Yet the present results suggest that these two processes are dissociable.

Figure Captions

Figure 1. Schematic description of the coherent motion task (a & b) and the 2-D form-from-motion task (c-f). See *methods* for description

Figure 2. Description of 3-D SFM stimulus construction for the face category. Static frames of the video were devoid of cues of shape. Once the stimulus was played and the object was viewed as rotating in depth, the structure of the object could be clearly identified using the 3-D SFM cues.

Figure 3. Performance of the SFM-impaired group. Although the performance of Patient 2 on the static shaded stimuli is worse than normal controls, it is still substantially above chance, while the performance of both patients on all SFM matching tasks is below chance. Scores are out of 15 for the naming task and out of 20 for the identification tasks.

Figure 4. Performance of the SFM-unimpaired group. Note that the performance of both patients is within 2 standard deviations of the normal control group, suggesting their performance to be similar to

controls. Scores are out of 15 for the naming task and out of 20 for the identification tasks.

Figure 5. Performance of the lesion-control group. The performance of the two lesion-control patients was similar to normal controls, within 2 standard deviation of the normal group’s performance. Scores are out of 15 for the naming task and out of 20 for the identification tasks.

Table 1. Details of the patients tested

Table 2. Coherent motion and 2-D form-from-motion thresholds for the patients tested. Units indicate % coherent motion.

Figure 1

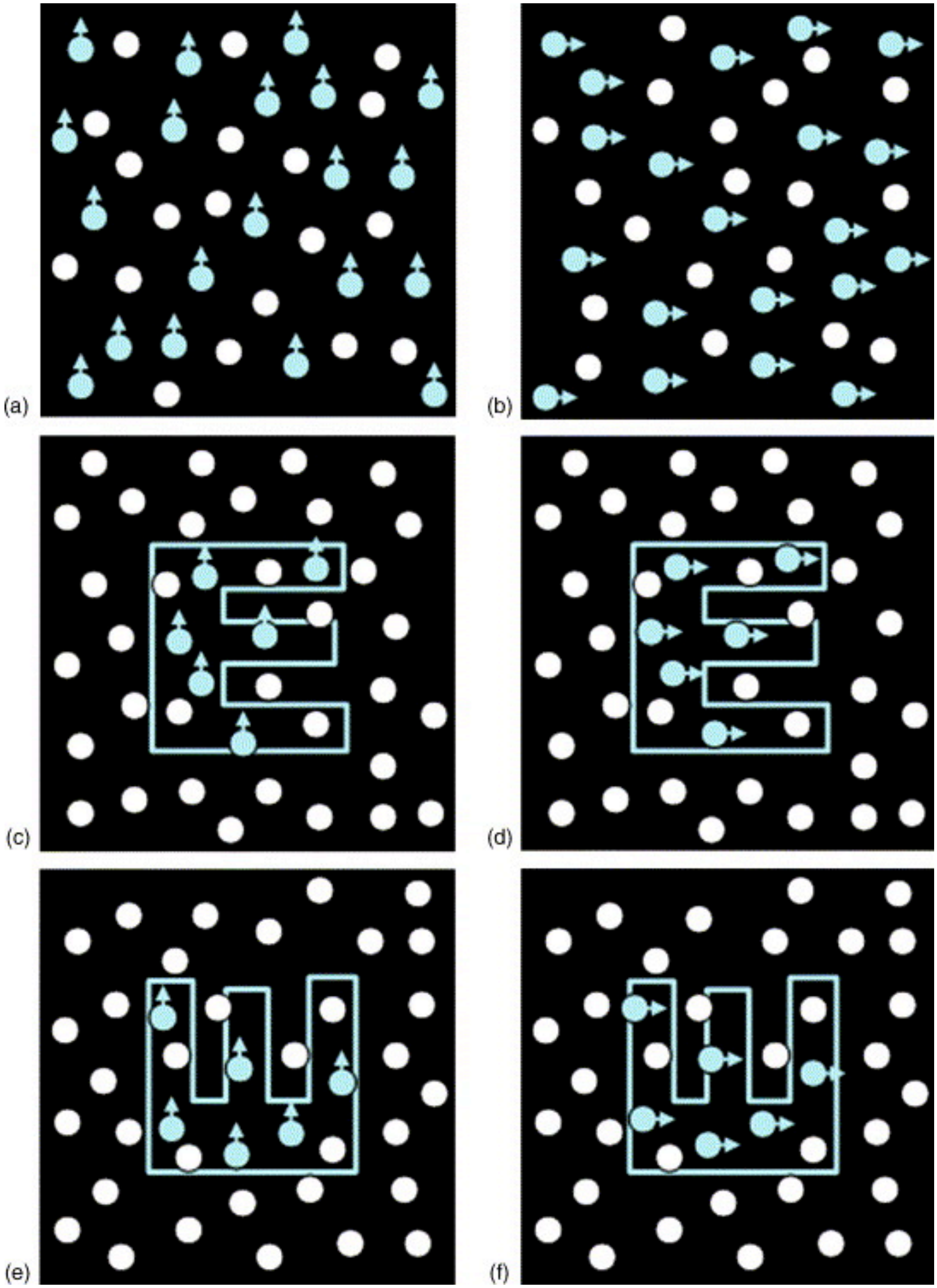


Figure 2

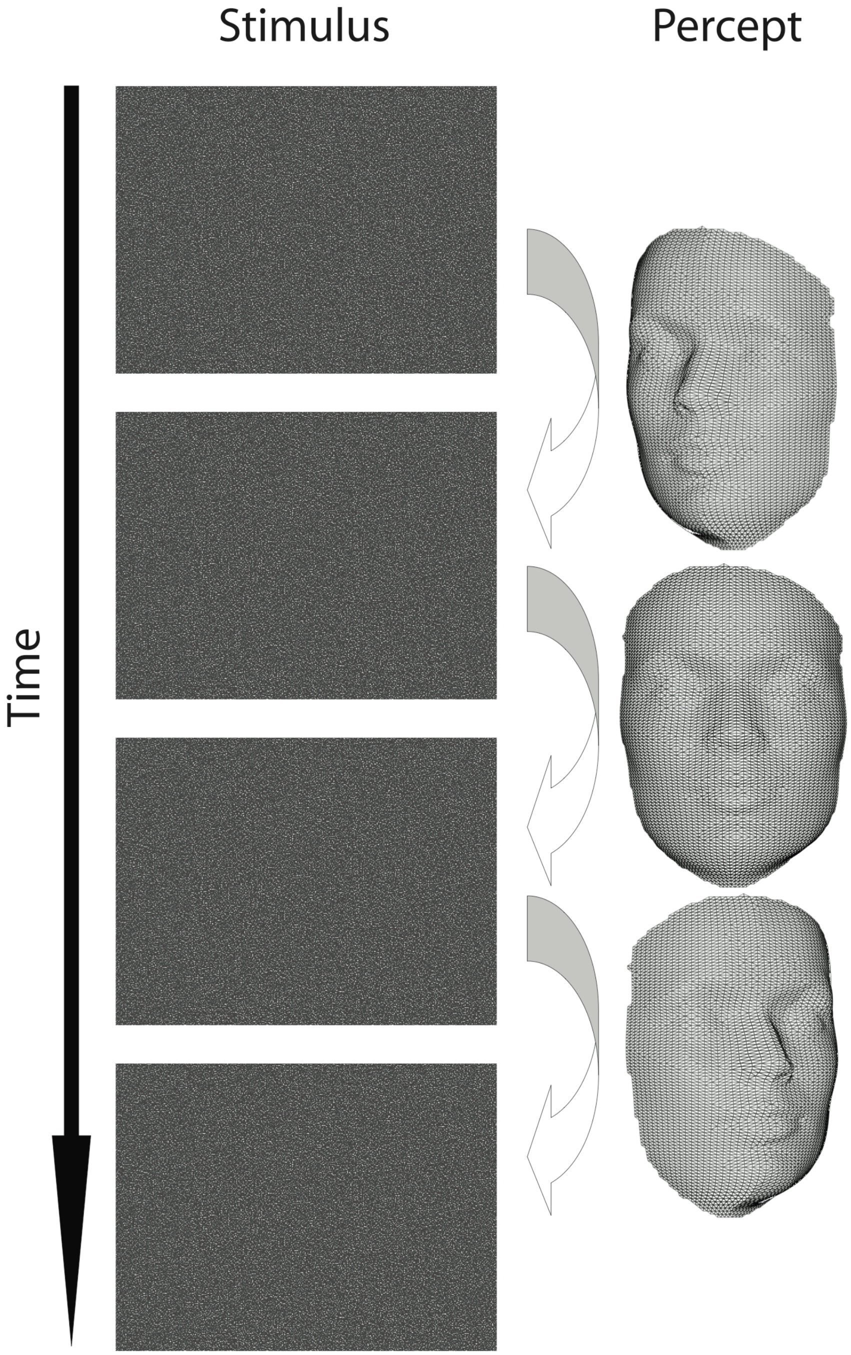


Figure 3

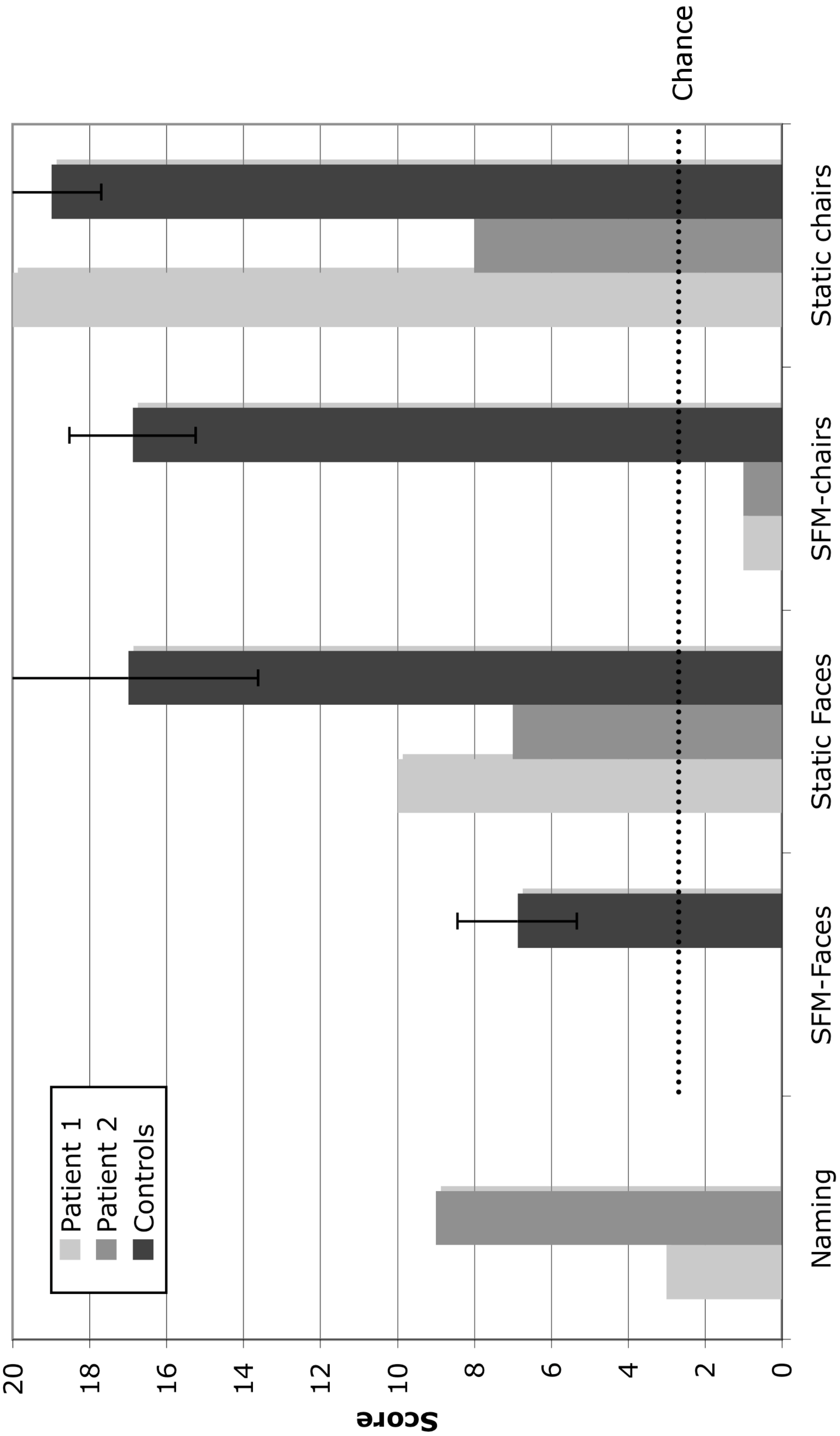

Score
0
2
4
6
8
10
12
14
16
18
20
Naming
SFM-Faces
Static Faces
SFM-chairs
Static chairs
Chance
Patient 1
Patient 2
Controls

Figure 4

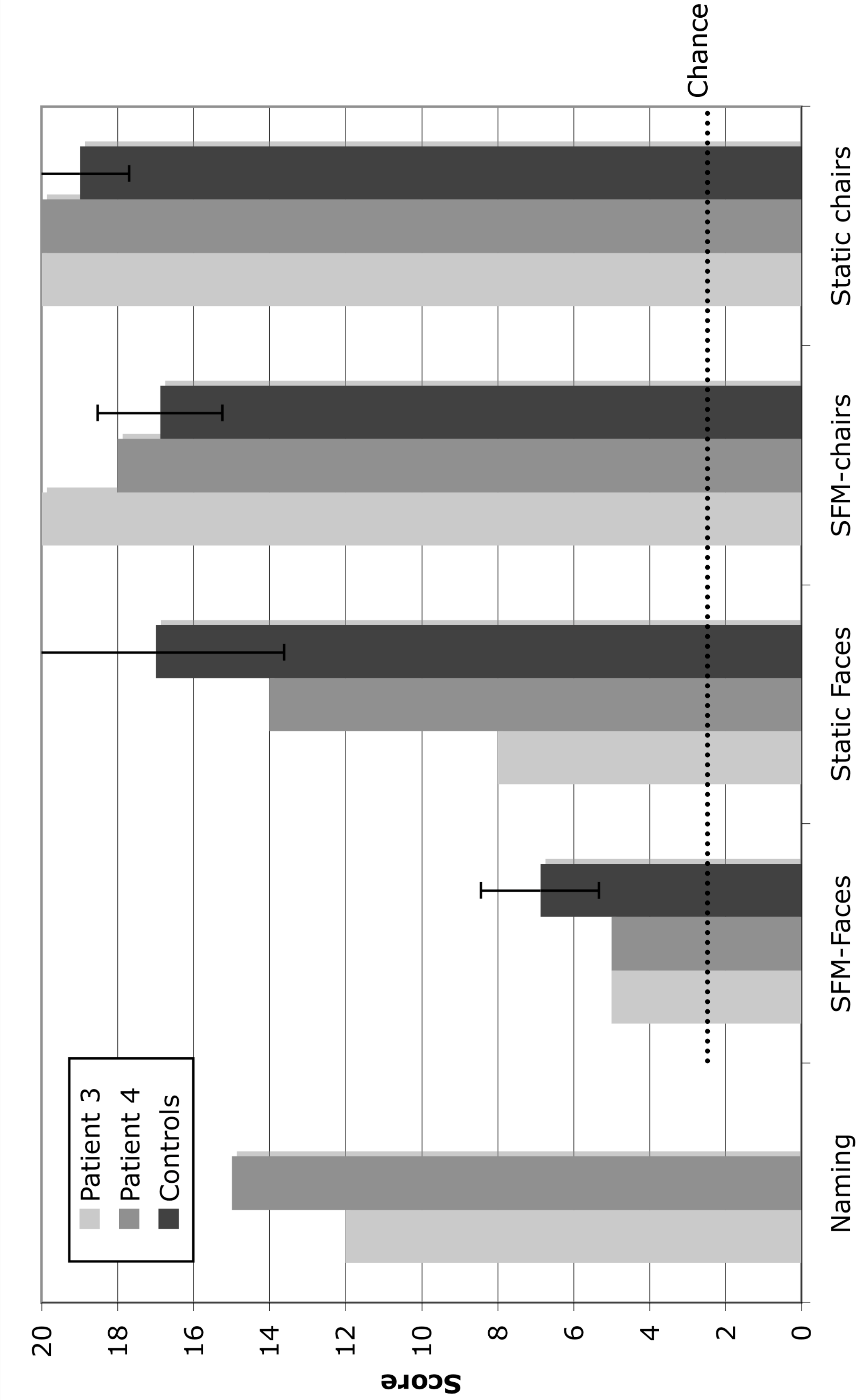

Chance
Patient 3
Patient 4
Controls
Score
20
18
16
14
12
10
8
6
4
2
0
Naming
SFM-Faces
Static Faces
SFM-chairs
Static chairs

Figure 5

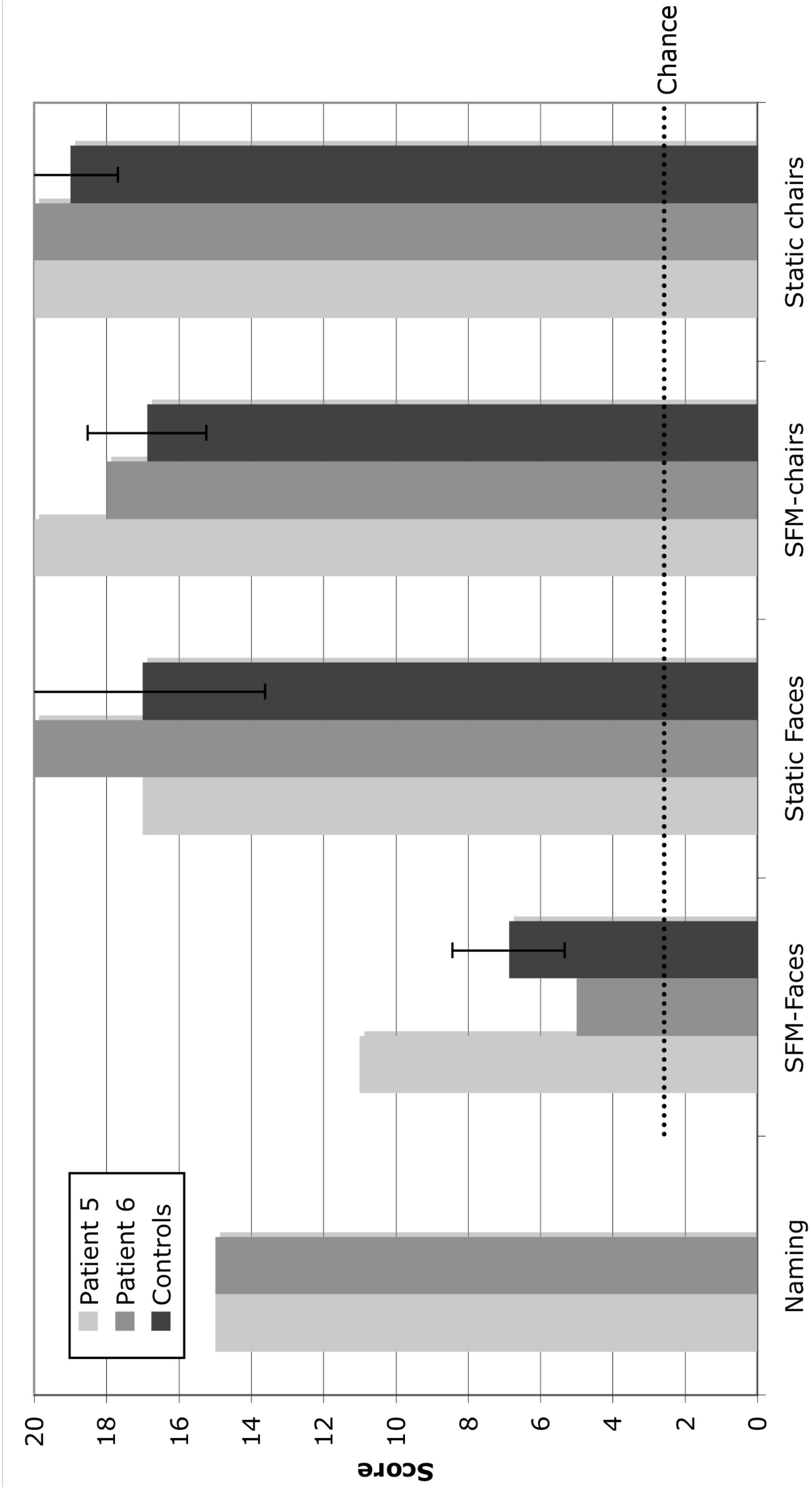

Score
0
2
4
6
8
10
12
14
16
18
20
Naming
SFM-Faces
Static Faces
SFM-chairs
Static chairs
Chance
Patient 5
Patient 6
Controls

Table 1

Subject	Gender	Age	Hemisphere	Lobe	Etiology	Visual fields
Patient 1	M	47	Bilateral	TP	Dementia	Full
Patient 2	F	72	Left	TP	Stroke	R, Inf quad.
Patient 3	F	65	Right	TO	Stroke	L scotoma
Patient 4	M	46	Right	TO	Malformation, Epilepsy	Full
Patient 5	M	21	Right	P	Malformation, Epilepsy	Full
Patient 6	M	63	Left	TP	Stroke	Full

TO—Temporal/Occipital; TP—Temporal/Parietal; P—Parietal;

Table 2

Patient	**Motion Coherence**	**2D FFM**
Patient 1	100	100
Patient 2	28.1	100
Patient 3	3.5	100
Patient 4	2.61	100
Patient 5	0.29	9.81
Patient 6	0.6	24.4

Chapter 5

CORTICAL REPRESENTATION OF COMPLEX OBJECTS DEFINED BY 3-D STRUCTURE-FROM-MOTION

The bulk of available evidence suggests that 3-D SFM cues are processed by dorsal stream mechanisms, while the identity of objects is determined by ventral stream mechanisms. Additionally it has been proposed that faces, as a highly homogenous and important category of objects, are processed by unique modules in the human brain, with an anatomical locus in the middle section of the fusiform gyrus. However, in addition to this face selective region—the so-called fFusiform Face Area (FFA)—a number of other regions have been found to be responsive more to faces than other objects. In the following paper, we assessed the role of the FFA and the Occipital Face Area (OFA) in the perception of 3-D SFM faces and chairs and to control scrambled stimuli.

Running Head: Motion-defined face processing

Cortical representation of complex objects defined by 3-D structure-from-motion

Reza Farivar, Michael Petrides, and Avi Chaudhuri

McGill University, Montreal, Canada

Correspondence to:

Reza Farivar
Dept. of Psychology
McGill University
1205 Doctor Penfield Ave
Montreal, Quebec, H3A 1B1
Reza.farivar@mail.mcgill.ca
www.cvl.mcgill.ca

Abstract

The rotation of objects in depth can evoke a strong perception of 3-D structure, a phenomenon termed structure-from-motion (SFM). This motion-dependent cue of depth is thought to be computed largely by mechanisms in the dorsal visual pathway that include areas MT, MST, STP, and LIP in the monkey brain. In contrast, visual areas in the ventral pathway, such as V4, TEO, and TE are thought to be essential for object recognition. Thus the recognition of objects from 3-D SFM cues is thought to be an example of dorsal-ventral integration, as postulated by a model of face recognition proposed by O’Toole et al. (2003). Using event-related FMRI in a passive viewing task in humans, we investigated the response of the occipital and mid-fusiform face-selective regions (OFA and FFA) to motion-defined faces, chairs, and scrambled versions of the stimuli. We found that only the right OFA retained category selectivity with 3-D SFM faces, while the right FFA showed a differential response for whole objects versus scrambled control stimuli. These results suggest that the OFA may be more sensitive than the FFA to facial surfaces defined by a 3-D cue, in contrast to a previous report by Kriegeskorte et al. (2003).

Introduction

The visual scene contains a wealth of cues for identifying its component objects. Whereas static depth cues such as shading and size can provide an understanding of an object’s shape in space, dynamic depth cues, such as 3-D structure-from-motion (SFM) can be also informative. Dynamic cues such as SFM are thought to be extracted by a pathway that begins in the primary visual cortex and extends dorsally to the posterior parietal cortex (Andersen & Bradley, 1998; Orban et al., 2005), whereas static shape cues are believed to be processed by a different pathway—one that extends ventrally to the inferior temporal lobe (Reddy & Kanwisher, 2006).

The evidence for this distinction comes from studies of patients with cortical lesions. Lesions of the inferior temporal lobe often cause impairment in recognition or memory, sparing spatial coordination and perception of motion (Damasio et al., 1982; Mishkin, 1954; Mishkin & Pribram, 1954), whereas lesions of dorsal stream areas such as the human homologue V5/MT or posterior parietal cortex often cause deficits in motion perception and spatial orientation or visually-guided actions, but leave recognition and memory intact (Goodale, Milner, Jakobson, & Carey, 1991; Holmes & Horax, 1919; Newsome & Pare, 1988; Pohl, 1973; Ungerleider & Mishkin, 1982). These patterns of behavioural deficits are consistent with anatomical studies in non-human primate models, which suggest that two separate anatomical pathways distribute information from early visual areas in a parallel and hierarchical manner.

Natural visual processing must involve the integrative action of both pathways. Connectivity studies in the macaque suggest that the two pathways, although largely distinct, do have sparse cross-connections before converging in the superior temporal polysensory area (STP; Jones & Powell, 1970; Seltzer & Pandya, 1978; Young, 1992). It has been postulated that such connections are necessary to solve the binding problem, although it is also possible that they may be involved in the transmission of shape information extracted from motion (Anderson & Siegel, 2005; Orban et al., 2005). In the current study, we explore the latter possibility.

Structure-from-motion cues are interesting because they are believed to represent a more complex form of motion processing, requiring upstream areas such as area MT for its computation (Andersen & Bradley, 1998). The responses of area MT neurons correlate with the perceived direction of the movement of a bi-stable, transparent, dotted cylinder rotating in depth (Andersen & Bradley, 1998). Cells in area MST show selectivity to rotation and tilt of depth-rotating planes in the absence of other cues (Sugihara et al., 2002). Selectivity to SFM attributes is also seen at the point of convergence between the two streams, i.e., area STP (Anderson & Siegel, 2005). Together, there is considerable evidence to support the role of higher visual areas in the extraction of surfaces from SFM. However, it is not clear whether these higher dorsal areas are sufficient for the perception and recognition of 3-D SFM surface. The case of patient D.F. (Dijkerman et al., 1999), who suffered damage to the ventral stream due to carbon monoxide poisoning leaving her incapable of recognizing or reporting form information about objects, suggests that perception of dynamic depth cues can occur without the involvement of the ventral stream. Although she is unable to use static pictorial depth cues, she can make use of dynamic information in a delayed reaching task. These results are highly suggestive of dissociation between the perception of SFM, which may be independent of the ventral stream, and the recognition of SFM-defined objects, which is dependent on dorsal-ventral integration.

We have developed a protocol by which we can assess the nature of dorsal-ventral integration. The stimuli consisted of depth-rotating laser-scanned facial surfaces with textures consisting of uniformly placed random dots. The faces are invisible without motion—once rigid head motion is made in depth, vivid details of the face become apparent. To assess the possibility of dorsal-ventral integration, we also tested two patients, one suffering from akinetopsia and another from prosopagnosia. We have found that whereas dorsal stream mechanisms are necessary for the perception of SFM faces, ventral stream areas are necessary for their recognition. In the study we report here, we attempt by way of a functional neuroimaging study in humans to test the involvement of the putative face-processing mechanisms in the perception of faces defined by motion.

An important factor driving our decision to use face stimuli was that this topic has received a great deal of attention, largely because of the social importance of face recognition, but also because of the intriguing idea that faces are neurophysiologically special. Functional imaging studies have repeatedly highlighted the involvement of the middle section of the fusiform gyrus in the perception of faces or face-like stimuli such as cartoons (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). Kriegeskorte et al (2003) assessed the processing of SFM-defined faces in an event-related FMRI study that compared the cortical response to two SFM-defined faces against random surface controls in a face-detection task. They reported selective engagement of the fusiform face area (FFA) for faces as compared to random surfaces, though they also reported elevated response in the FFA for the type of stimuli that subjects found more difficult to perceive as faces—novel 3-D motion defined stimuli that are formed by the movement of dots over a surface, akin to a series of laser beams scrolling over a surface. Three aspects of the Kriegeskorte et al. (2003) study require further consideration.

First, the face stimuli used were not strong enough to elicit a vivid face percept for naïve subjects and subjects required familiarization to distinguish between the face stimuli and random surfaces. Second, a comparison to other meaningful objects was not carried out—the cortical response to faces was compared to random surfaces. This biases the results in that the faces were the only meaningful complex stimulus used and it is thus impossible to assess whether the FFA was selective to SFM-defined faces or simply responsive to them. Third, the fact that subjects showed elevated responses to their novel on-surface stimuli as compared to classic SFM-defined faces in the FFA is difficult to appreciate. They report that subjects found these face stimuli harder to perceive, thus as less face-like, yet the FFA was engaged more strongly for these hard-to-perceive faces than the 3-D SFM faces. One possible explanation is that the subjects had to engage in visual imagery, more so for the difficult stimuli than the easier ones, and this resulted in the greater response of the FFA (O'Craven & Kanwisher, 2000).

Methods

Stimuli

The face stimuli consisted of 40 laser-scanned facial surfaces taken from the Max-Planck Face Database (Troje & Bulthoff, 1996). The surfaces were rendered with a volumetric texture map that ensures uniform texture density across the surface—it is analogous to carving a surface out of a stone block. Shadows and shading were removed from the rendering, as previously described (Liu et al., 2005). The faces were rendered against a similarly textured random-dot background. During the animation, the face rotated from -22.5 degrees to 22.5 degrees, centered at the frontal plan, in one cycle. This rotation was captured in a 100 frame video that lasted 3.3 seconds.

The 40 chair stimuli were obtained from chair model databases. They were rendered in exactly the same manner as the faces. Scrambled 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 scrambling their positions. The resulting scrambled stimuli share many of the low level features of the original videos but do not carry any object identity or meaning.

For the FFA localizers, a series of 140 frontal face photos were used in addition to 140 house photos. The images were gray-scale and were of equal average luminance.

Subjects

Four male and six female subjects (mean age = 24 years, range: 20-29) participated in the study. Informed, written consent was obtained from all of the participants according to the institutional guidelines established by the Ethics Committee of the Montreal Neurological Hospital and Institute.

Experimental Setup

The stimuli were rear-projected onto a translucent screen that was reflected off a mirror mounted above the subjects' face. The image subtended approximately 10 degrees of visual angle vertically and approximately 13 degrees horizontally. Subjects reported facility at vividly perceiving all the stimuli within this setup. The subjects' eyes were monitored at all times with an MR-compatible camera (MRC Systems GmbH) that was mounted on the head coil and maintained an unobstructed view of the eyes.

Acquisition

MR acquisition was performed on a 1.5 T Sonata MRI Scanner (Siemens, Erlangen, Germany). After a high-resolution T1 anatomical scan (entire head, 1 mm^3 isotropic resolution), three runs totaling 596 images sensitive to the Blood Oxygenation Level Signal (BOLD) were acquired with the following settings: 38 oblique T2* gradient echo-planar images; voxel size, 3.4 x 3.4 x 3.4 mm; repetition time (TR), 3.5 s; echo time, 50 ms; flip angle, 90°. For the localizer, an additional 115 frames were acquired with the same acquisition parameters.

Procedure

All subjects readily perceived the SFM faces and chairs and thus no training was necessary. The main experimental trials were divided into 3 runs. Within each run, 20 examples of each stimulus category were randomly presented, for a total of 60 repetitions of each condition. The 3.3 second videos were followed by a fixation period of variable length, with the length of the fixation period being sampled from a Gaussian distribution with a mean of 5 seconds and a standard deviation of 1 second. The FFA localizer consisted of a block-design session alternating between 30 seconds of face images, each presented for 600 msecs with an ITI of 282 msecs, 15 seconds of rest with fixation, and 30 seconds of house images presented in the same manner as the faces, followed again by the 15 seconds of fixation. In total, four blocks of faces and four blocks of houses were presented. Throughout all the runs, subjects were asked to maintain fixation and view the stimuli attentively but passively. Their eyes were monitored at all times with the MRI-compatible camera.

Analysis

All preprocessing and statistical analyses were carried out with the SPM2 software, and ROI analysis was carried out with the MarsBar extension for SPM2. Slice-time correction with sinc interpolation was first carried out to temporally align the slices within a volume. The first three frames in each run were discarded to allow for equilibration effects, and the remaining frames across all the runs were realigned to the fourth frame in the first run and smoothed with a Gaussian kernel 8mm full-width at half-max. Low frequency drifts were removed with a high-pass filter with a period of 128 seconds, and serial autocorrelations were estimated as a first-order autoregressive process using the AR(1) method. Parametric statistical models were estimated at each voxel using GLM with a model of event times convolved with a canonical hemodynamic response function and its delay and dispersion derivatives. Using the high-resolution T1 image, a transformation of the individual subject brain to the MNI template was estimated and was then applied to the individual SPMs to allow for group analysis and aid interpretation. Random-effect group analysis was carried out by entering the contrast estimates from the individual subject analysis into a second-level between-subject analysis evaluated by a single-sample t-test against a mean of zero. The face-selective ROIs were defined by a 6mm radius sphere centered at the peak of the cluster in the mid-fusiform and lateral occipital gyri as estimated from the localizer run. The coordinates for the ROI centers are given in Table 1.

Results

ROI Analysis

For the ROI analysis, the realigned images prior to smoothing were used to avoid pooling of voxels neighbouring the ROI. To test the selectivity of this ROI for the processing of motion-defined faces, we carried out the same contrasts but within the specified ROIs. The contrast estimates for each subject were then entered into a second-level, random-effects analysis and evaluated with a single-sample t-test for each ROI. Figure 1 represents the estimates of percent signal change for the conditions across the ROIs, with error bars depicting between-subject error. It should be noted that the ROI analysis was not conducted using percent signal change values within each ROI (as reported in Figure 1), but rather by estimating the contrasts at each ROI and testing the average of the contrasts against a mean of 0 (no effect) using a single-sample t-test. This is the same method used for the global SPM analysis and therefore maintains a degree of consistency.

We did not observe a selective involvement of the mid-fusiform in the processing of faces over that of chairs, but interestingly, we found a significantly greater involvement of the right mid-fusiform for the processing of whole SFM stimuli compared to the scrambled control (t(8) = 3.349, p<0.005). There are at least two important points to consider in determining the source of the difference in the finding of our study compared to that of Kriegeskorte et al. (2003). First, the Kriegeskorte et al. (2003) study did not examine the response of FFA to faces versus other objects. Their control stimuli consisted of random perturbations in a surface, resulting in a 3-D surface that lacks meaning as a real object. Here, we used chairs as the comparison stimuli, because an important quality of the FFA is that it responds more strongly to faces than to other classes of objects. Contrasting the response of the FFA to face stimuli to its response to random shapes would not allow for an adequate validation of its selectivity. It may thus be the case that similar levels of activation were observed in the Kriegeskorte et al. (2003) study as that reported here, but because the response of the FFA to random patches would be minimal, they observed a significant involvement of the FFA in 3-D SFM perception whereas we did not.

Second, the stimuli used in the Kriegeskorte et al. (2003) study were difficult to perceive, necessitating familiarization for the subjects. It may be the case that the FFA was activated more because the subjects tried to see a face in the stimulus than actually perceiving the face vividly, as is the case with our stimuli. Previous reports have suggested that even imagery for faces can result in activation of FFA (O'Craven & Kanwisher, 2000). Kriegeskorte et al. (2003) used two categories of motion-defined stimuli—the classic 3-D SFM stimuli, and a novel SFM stimulus analogous to a face in the dark whose surface is exposed by the traveling incidental laser spots. Their subjects reported they found “on-surface SFM perception more difficult”, yet this stimulus type resulted in stronger activity in FFA than classic 3-D SFM stimuli. This finding supports the notion that subjects may have engaged the FFA because the difficulty of the task may have necessitated use of imagery to perceive correctly the stimuli.

Our results may at first appear at odds with the results of Kriegekorte et al. (2003) who reported increased activity in the FFA for SFM-defined faces, but in fact we replicate their findings. Note the contrast here is between whole and scrambled objects, which is effectively the same contrast they reported: SFM-faces compared to SFM random phase-scrambled surfaces.

Our stimuli were readily perceived as faces. The fact that these stimuli did preferentially activate other regions, such as the inferior portion of the lateral occipital gyrus, roughly corresponding to a region ventral of the LOC, suggests that these stimuli did stimulate object recognition mechanisms, but not the FFA. Thus it cannot be argued that our stimuli were too “weak” to stimulate the FFA, because they were strong enough to elicit activity in other high-level object regions.

The same analysis described above was carried out for the region in the lateral portion of the inferior temporal gyrus that is more sensitive to the presentation of faces than to other objects, often termed the Occipital Face Area (OFA). This area is reported to be less reliably identified as face-selective than the mid-fusiform area (Kanwisher & Yovel, 2006), but activity in this region correlates well with the perception and recognition of faces (Grill-Spector, Knouf, & Kanwisher, 2004). Moreover, this region may be an integral part of a larger, distributed face-processing network (Haxby, Hoffman, & Gobbini, 2000; Rossion et al., 2003).

The random-effects analysis of the right occipital face-selective region suggested a significantly greater engagement of this area for motion-defined faces than motion-defined chairs ($t(7) = 1.92$, $p<0.05$). None of the other contrasts were statistically significant in the random-effects analysis for this area. For the contrast of whole objects versus scrambled controls, only the right FFA showed selective activity. Also, as depicted in Figure 1, the OFA on both hemispheres appear to be more responsive to faces than all other stimuli.

Global SPM Results

Three contrasts were used to evaluate the cortical response to SFM faces and chairs. We evaluated the simple effects of Faces versus Chairs and Scrambled Faces versus Scrambled Chairs. Our hypothesis was that face-selective regions would exhibit a significant difference in their response to faces as opposed to chairs, but would fail to show the same difference when comparing scrambled faces to scrambled chairs. In addition, peaks were considered face-selective if they survived a larger contrast comparing response of faces to all other objects. These criteria together would better ensure that a putative face-specific region was not responding to simple curvature information in the face or any low-level difference between the face and chair stimuli and that is actually face-specific. In addition to the analyses of these simple effects, we evaluated the differences in the cortical response to whole and scrambled objects. Given that, with the exception of Kriegeskorte et al. (2003), our study is amongst the first to evaluate the cortical response to meaningful SFM stimuli, we wondered whether these stimuli would activate the same ventral stream regions that photographs of objects do.

Figure 2 depicts the results of these contrasts represented on an average fiducial map thresholded at p < 0.005 (uncorrected), with average delineations of visual areas as obtained from published results

(Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998). Structure-from-motion faces affected the response of the right lateral occipital face region (OFA) more than chairs. This region was more ventral than the published locus of area LOC. This pattern was observed in nine subjects, whereas the right FFA of only one subject was found to be selectively engaged by the SFM faces. We did not find evidence for the engagement of the OFA for the control stimuli (scrambled faces versus scrambled chairs), suggesting further that this region was indeed face selective and not simply selective for low-level properties of the face stimuli.

Discussion

Our main finding is that motion-defined faces selectively engage the right lateral occipital face-responsive region in the absence of a similar pattern of activity in the right mid-fusiform gyrus. The lack of a face-selective response in the mid-fusiform is perhaps the more surprising finding here, because this region has been repeatedly shown to be highly selective to the perception and recognition of face images, yet in our study with readily-perceivable motion-defined faces, we were unable to detect a selective engagement of this region for faces.

Lack of a selective response in the mid-fusiform to faces

A number of explanations can be offered for this null finding. First, it could be argued that the event-related design was not sensitive enough to capture the engagement of the mid-fusiform. Yet activity was observed in the mid-fusiform, it simply did not dissociate between SFM faces and non-face stimuli, thus it cannot be held that the stimuli simply did not activate this region. Second, a claim may be made that the stimuli did not elicit a strong-enough face percept to bring about selective engagement of the FFA. Yet, the same stimulus did elicit a selective response in the lateral occipital face region, and all subjects readily reported perceiving a face in the stimuli without any training. Thus, both through behavioural and neural evidence we can assume that the stimuli were indeed strong enough to elicit a selective response. Third, given that attention was not specifically engaged, it may be that differences in attention as relating to the different stimuli may have brought about the null results. One would normally expect faces to engage attention more strongly, as we have previously shown (Borrmann, Boutet, & Chaudhuri, 2003; Langton, Law, Burton, & Schweinberger, 2007) and therefore if attention capture differed between the stimuli, then one would expect more attention to the faces and an enhanced response to that category. Fourth, given the passive nature of the stimulation in this study, one may view the results as limiting because the subjects may not have attended to the stimuli altogether. This is highly unlikely because we recorded the eyes using the MRI-compatible camera and verified all the subjects observed the stimuli at all times. Furthermore, if the stimuli had not been attended to at all, we would have expected the null finding over the entire span of the cortex. Yet, we did observe the selective response in the lateral occipital and other regions, suggesting the subjects did follow the instructions and attended to the stimuli. The most parsimonious explanation is that the study by Kriegeskorte et al. (2003) report FFA activity to SFM faces because they compared response to face stimuli to the response to random surfaces, while ours compared face stimuli to other meaningful stimuli. Note that we were able to replicate the findings by Kriegeskorte et al. (2003), namely an enhanced response in the FFA for whole objects than scrambled ones, but we did not observe a selective engagement of FFA for faces when compared to chairs.

The role of the right lateral occipital region in face perception

It is known that lesions of the ventral aspects of the lateral occipital cortex can disrupt face recognition without damage to the mid-fusiform cortex (Rossion et al., 2003). Our data suggest that this region is heavily engaged in 3-D facial forms as opposed to other objects. We may thus speculate that the OFA plays an important role in the extraction of facial 3-D information for recognition. We cannot wholly discount the possibility that this region is simply more responsive to 3-D curved surfaces (as our face stimuli were mostly curved), but the fact that the response to the scrambled faces remains low despite an equal amount of surface depth and curvature suggests that this region is not simply selective to curved surfaces.

Although we did not specifically map area LOC in each subject for an ROI analysis, we did not observe selective engagement of this region in the group average data for whole objects as compared to their scrambled counterparts. This finding is surprising considering that this region was originally mapped by comparing images of whole objects with scrambled versions of the same images. It is admittedly difficult to reconcile these differences, but one possible explanation may be that area LOC is selective for plausible surfaces. This is an important quality of a natural scene photograph that is disrupted by scrambling. However, our method of scrambling, while destroying the object meaning, still retains a plausible object surface, much like an abstract sculpture. In other words, scrambling a photo disrupts the arrangement of surfaces in the objects present in the scene such that it becomes physically impossible to ever encounter a scene resembling the scrambled scene. Yet, the scrambling used in our study does not challenge physical reality, and we speculate that this plausibility of a scene is what the LOC may be sensitive to.

Conclusions

Structure-from-motion, as a cue to shape, allows for vivid percepts of complex objects including faces. The cortical representation of shapes defined by this 3-D cue appears to be different from the cortical representation formed from real images of the objects. Specifically, high-level face selectivity observed in the middle section of the fusiform gyrus is absent in the contrast of SFM faces with SFM chairs, but is present in the occipital face region, suggesting this region may play a greater role in the formation of a 3-D face percept.

Figure Legends

Table 1. Centers of the Regions of Interest in MNI space for all subjects (in mm).

Figure 1. Mean percent signal change across conditions and ROIs. The error bars reflect the standard error of the mean (between-subjects).

Figure 2. Global SPM results for contrast of (a) faces vs. chairs, (b) scrambled faces vs. scrambled chairs, and (c) whole vs. scrambled. t-maps are thresholded at $p<0.005$ (uncorrected).

Table 1

	Left FFA			Right FFA			Left OFA			Right OFA		
	x	y	z	x	y	z	x	y	z	x	y	z
SS1	-40	-58	-20	48	-52	-16	*	*	*	52	-76	-6
SS2	-42	-52	-20	42	-50	-22	-38	-88	-10	54	-74	-10
SS3	-44	-50	-20	38	-70	-12	-38	-90	-2	28	-94	-4
SS4	*	*	*	38	-66	-18	-26	-106	-8	56	-74	0
SS5	*	*	*	*	*	*	*	*	*	54	-74	2
SS6	-50	-64	-26	46	-50	-26	*	*	*	*	*	*
SS7	-38	-50	-20	44	-58	-20	-36	-80	-18	*	*	*
SS8	-48	-60	-24	46	-60	-22	-48	-73	-14	48	-75	-6
SS9	-40	-62	-16	42	-68	-12	-42	-86	-12	46	-75	-6
SS10	-43	-57	-23	42	-68	-12	-44	-78	-11	43	-79	-5
Mean	-43.1	-56.6	-21.1	42.9	-60.2	-17.8	-38.9	-85.9	-10.7	47.6	-77.6	-4.4
SD	4.1	5.4	3.1	3.5	8.2	5.1	7.0	10.7	5.0	9.1	6.8	3.8

Figure 1

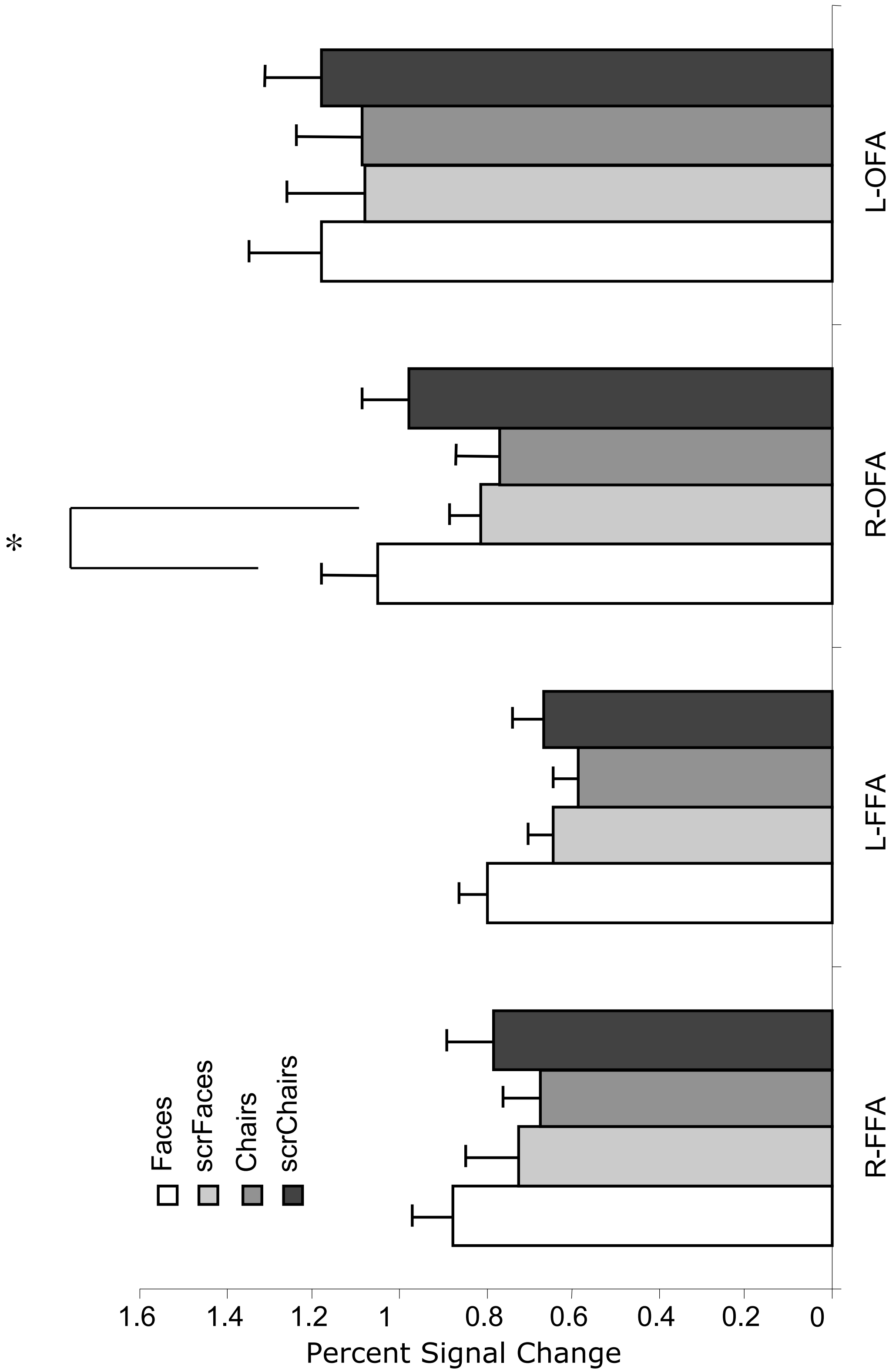
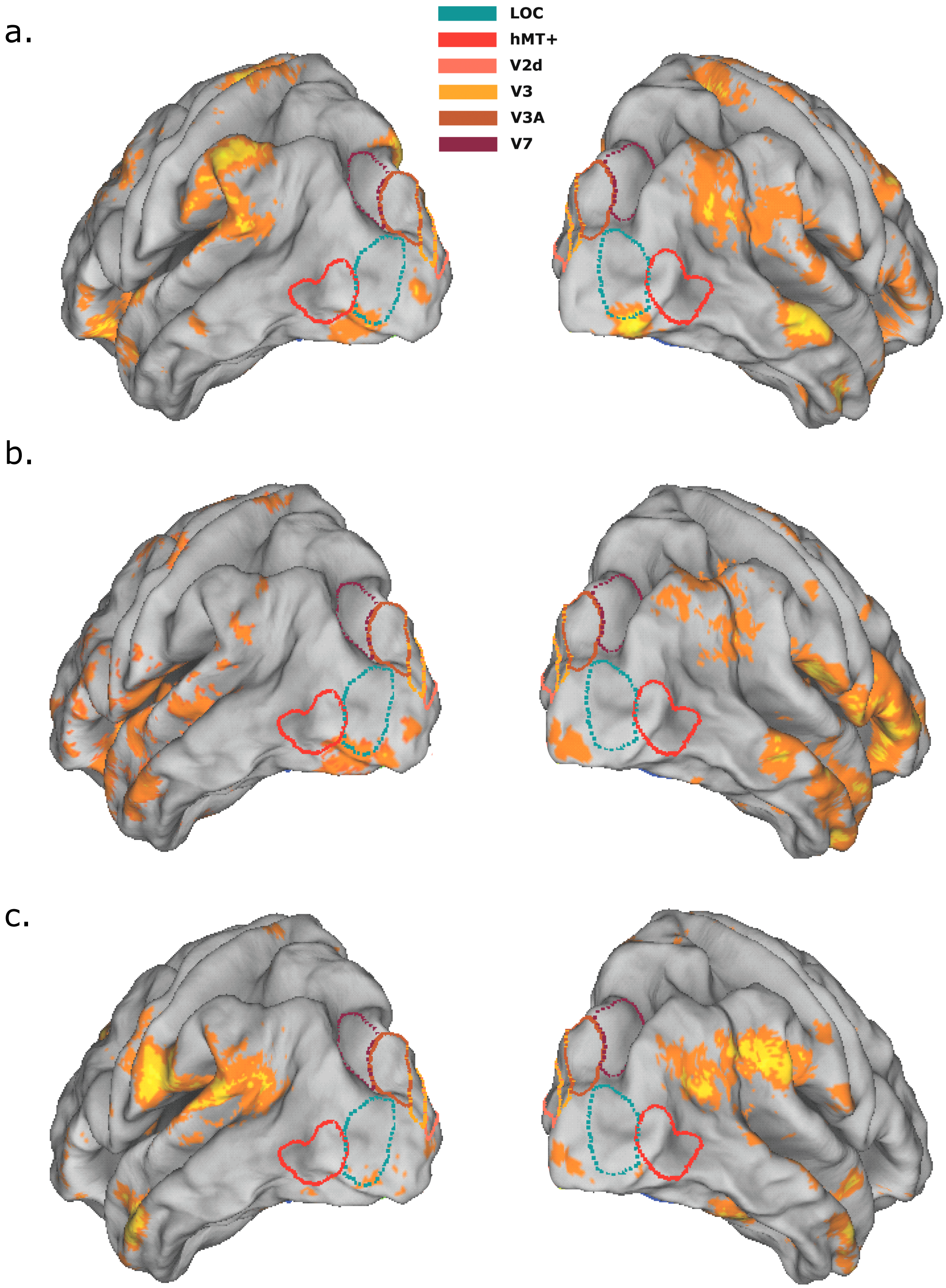

*
□ Faces
▨ scrFaces
▨ Chairs
■ scrChairs
L-OFA
R-OFA
L-FFA
R-FFA
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Percent Signal Change

Figure 2



Chapter 6

CONCLUDING REMARKS

In this chapter, the general results of the thesis chapters are summarized. Following a discussion of the work and the assumptions made, the future implications of this research will be considered in a broad scope.

Summary

Chapter 2 provided a synthesis of a diverse body of work that together may aid us in better understanding the distribution of labour in the cortex for complex object recognition. It was hypothesized that dorsal stream mechanisms responsible for the extraction of 3-D information from cues such as SFM forwarded their input to ventral stream mechanisms that represented objects in a viewpoint-invariant manner, given the fact that 3-D representations are quasi viewpoint-invariant. In this manner, the processes of normal object recognition can be seen as distributed between both the ventral and the dorsal visual pathways. The fact that familiarity may give rise to viewpoint-invariance (Booth and Rolls, 1998) and expertise is derived from extensive familiarity together suggest that objects with which we have developed some expertise, such as faces, may have the potential for being more viewpoint-invariant than non-expert objects. Evidence in support of this view was discussed.

Chapter 3 utilized modern computer rendering methods to generate complex 3-D shapes that are defined solely by SFM cues and are devoid of other cues of shape such as shading. This development allowed us to directly evaluate the model proposed by O'Toole et al. (2002) and also to attempt to replicate the earlier experiments of Bruce and Valentine (1988). Ours was the first such direct report of SFM cues driving complex recognition and as such supported the view stated in Chapter 2 that cues of shape that are understood to be dorsally computed can nonetheless empower ventral stream mechanisms.

In Chapter 3, we also evaluated the hypothesis that the recognition of SFM faces is actually driven by ventral stream processes, while the extraction of the surface detail may have been dorsally computed. By evaluating a patient with akinetopsia—motion blindness—we were able to confirm that intact dorsal stream mechanisms are necessary for the extraction of SFM surfaces from depth. Testing face recognition capacity in a previously reported prosopagnosic patient (patient P.S., reported by Rossion et al., 2003) allowed us to test whether or not recognition of unfamiliar SFM faces requires the ventral stream. We found this to be the case. Taken together the findings of Chapter 3 suggested that (a) SFM cues can drive unfamiliar face recognition, and (b) this paradigm may represent a clear example of dorsal-ventral integration.

Chapter 4 evaluated the extent to which complex motion perception processes can be considered unitary. An implicit assumption has often been that perceptual processes can be subdivided into lower and higher level echelons, with 2-D processes taking the lower ranks and 3-D processing the upper. This notion implies that damaging 2-D processing should also render the mechanisms of 3-D computations impaired. Vaina et al. (1990) tested this hypothesis initially with a transparent rolling cylinder task that required subjects to evaluate the direction of rotation of a transparent rolling drum. While the task is indeed complex, it is unclear whether it actually tested 3-D SFM perception or not, because one could succeed or fail on the task simply on the basis of their capacity for processing motion transparency. A valid examination of 3-D SFM perception requires one to directly and specifically test objects whose internal structures are visible only if one is able to extract a 3-D surface from SFM cues. The tasks we used did not rely on motion transparency and indeed, required the extraction of the internal features of the 3-D surfaces and thus served as good tests of SFM capacity. In a group of patients all with elevated 2-D form-from-motion thresholds, we found that some could perform well above chance on the 3-D SFM task, suggesting that 2-D FFM performance does not predict 3-D SFM performance. As such, the processes of complex motion may be dissociable from one another.

The functional imaging study reported in Chapter 5 evaluated the hypothesis of dorsal-ventral integration for SFM faces by comparing cortical response to SFM faces and chairs in contrast to scrambled versions of each. We asked whether the middle section of the fusiform gyrus, termed the fusiform face area (FFA) by Kanwisher et al. (1997) is indeed sensitive to 3-D SFM cues. If so, then this area would be more strongly engaged by 3-D faces defined by SFM than by 3-D chairs defined by SFM, and more so to both than to scrambled shapes that lack meaning. We also evaluated another cortical area considered important for face recognition, the occipital face area (OFA) using ROI analysis. Our results suggest that the FFA is not selectively engaged by the 3-D SFM faces, but the OFA is. Interestingly, patient P.S., suffering from prosopagnosia, lacks the right OFA but has a normally responsive right FFA. We interpreted our data to suggest that the OFA is more important for extraction of 3-D facial surfaces than the FFA, and it is for this reason that our stimuli tapped into this area more strongly.

The data reported in Chapter 5 stand at odds with a related study by Kriegekorte et al. (2003), but as discussed in Chapter 5, the Kriegeskorte et al. (2003) study did not directly compare the response of the FFA to faces and another object category as we have—they compared the FFA response to SFM faces to an SFM random surface. This is analogous to the comparison of the cortical response to whole versus scrambled stimuli, for which we observed a significant contrast in the FFA.

In summary, the results reported in the dissertation suggest that (a) SFM cues are sufficient to drive complex object recognition, such as unfamiliar face recognition, (b) putative dorsal stream mechanisms are necessary for the extraction of surfaces from motion, (c) putative ventral stream mechanisms are necessary for the recognition of objects defined by SFM, (d) 3-D SFM and 2-D FFM may be dissociable processes, (e) SFM-defined faces engage the occipital face area more than SFM-defined chairs, and (f) the FFA may be less sensitive to discrimination of objects defined by SFM cues.

Assumptions and Limitations

The results reported and interpreted in this dissertation are all based on the assumption that human cortical organization mimics that of macaques—the main primate model for which we have extensive connectivity data. There is no guarantee that this is the case. Indeed, we have to date no direct connectivity data on the organization of the human visual cortex and thus cannot confirm that the same pattern of connectivity and limits on connectivity that exist in the macaque brain also exist in the human brain. It may be the case that dorsal and ventral pathways, generally accepted to exist in the human brain as well, have more cross connections than observed in the monkey brain, in which case there may be more opportunities for the two pathways to interlink and communicate.

Using FMRI, it has been possible to at least compare the degree to which the pattern of activity seen in the human and monkey brain to the same stimuli correspond. Vanduffel et al. (2002) for example report high degree of homology between the two species, with differences as well. Approaches such as functional imaging in awake primates and its comparison to humans may enable us to more directly compare functional homologies between the two species and thus increase our confidence in interpreting human data based on cortical organization models obtained from the monkey brain.

Another important assumption made in the interpretations reported in this dissertation is the locus of SFM processing in the human brain. It is generally assumed that visual motion signals are processed by dorsal-stream mechanisms, and it is true that no data exists to suggest that complex motion perception, such as SFM, is computed in ventral stream areas. The bulk of neuroimaging data on humans suggests that dorsal stream mechanisms are necessary for the extraction of SFM surfaces, but ventral stream areas are also sensitive to shapes defined by SFM. It is unknown whether ventral shape representations communicate back with their dorsal counterparts, perhaps to stabilize or interpret the SFM-derived shapes. Thus the mode of interpretation here assumes that dorsal-ventral integration for the recognition of complex objects is unidirectional—that dorsal stream areas provide input to ventral stream object recognition processes. While this view is consistent with current models of objects recognition, connectivity data on the macaque brain does suggest that regions in the IT do send inputs to dorsal stream areas as well (Webster, Bachevalier, & Ungerleider, 1994). Thus there may be a role for ventral-to-dorsal input that is not investigated here nor taken into consideration in interpreting the results.

Another possibility for information transfer is via the feedback connections to earlier visual areas from higher areas. For example, while cells in MT and upstream exhibit increasingly greater selectivity for 3-D cues such as SFM, these regions also provide feedback connections to earlier areas, including V1. Thus it is unclear whether the dorsal-ventral integration postulated here is achieved by direct connections at the high levels of the two hierarchies or via feedback to earlier areas and subsequent forwarding. The results of Grunewald et al. (2002) suggest that V1 neuronal responses do not correlate as well as MT cells with the perceived direction of rotation for a rotating transparent cylinder, but they do report V1 activity that correlated with the perception nonetheless. The authors suggested that the pattern exhibited in V1 may have been shaped by the top-down input from MT. Thus the possibility exists that if the extraction of SFM cues is restricted to the dorsal stream, the forwarding of this information to relevant regions in the ventral stream could go via two pathways, a direct and an indirect one.

Another important assumption made here is that the SFM-defined faces are perceived and understood in the same manner that one understands a normal face photograph or a real face. The stimuli used for the experiments reported here are more akin to facial masks than to real faces. However, the fact that such faces cannot be recognized by a prosopagnosic patient suggests that whatever mechanisms are engaged by the SFM-defined faces are akin to the same processes engaged by face photographs or real faces. It is unclear whether the cortical representation of SFM faces is identical to that of face photographs, but here it is generally assumed to be the case and we did not observe data to the contrary in the FMRI study reported, with the exception of a lack of FFA selectivity for SFM-defined faces. In addition, we have further evidence, not reported in this dissertation, that certain effects seen with face photographs, such as the composite face effect (A. W. Young, Hellawell, & Hay, 1987) may also be seen with SFM-defined faces, and that the electrophysiological correlate of human face perception, the N170 face-specific component (Bentin, Allison, Puce, Perez, & McCarthy, 1996), may also be present for these SFM-defined faces. Taken together, it is more parsimonious that SFM-defined faces do tap into the same mechanisms that are engaged during normal face perception and recognition.

Related phenomena and future work

There is surprisingly little data on the effects of dorsal stream stimulation or inhibition and ventral stream recognition. The data presented in Chapter 3 suggest that dorsal stream mechanisms are important for the extraction of surfaces, but it would be pertinent to demonstrate this in normal subjects using temporary deactivations by means of transcranial magnetic stimulation (TMS). It is known that TMS of hMT+ may disrupt motion perception in a manner resembling results from microstimulation of monkey V5/MT. Thus it would seem plausible that stimulation of hMT+ or the 3-D selective IPS regions should disrupt successful perception of 3-D shapes defined by specific cues such as SFM.

The pattern of dorsal-ventral integration described herein is assumed to proceed in a serial manner, with dorsal stream regions first extracting the surfaces from motion and ventral stream regions receiving and processing this input for recognition. If true, then one would expect the time course of activity in the dorsal stream as measured electrophysiologically to precede selective activity in the ventral pathway. For example, one would expect that a dorsally-originating ERP signal that is selective for coherent motion or structure-from-motion to precede the face selective N170 component when a subject views SFM-defined faces. This is currently being investigated in collaboration with Miguel Castelo-Branco and Peter de Weerd.

An issue related to dorsal-ventral integration is cue-integration and competition. We assume, for example, that static cues of shape, such as texture or shading, may be processed via the ventral pathway and may not require involvement of the dorsal pathway. If so, then how do dorsally-originating cues such as SFM and perhaps stereopsis interact with ventrally-originating cues, such as shading? Do the two combine prior to the evaluation of objects, or do we have multiple cue representation of objects? If in conflict, do they compete, resulting in a bi-stable percept, or does one overwhelm the other?

A set of experiments are currently underway to address these questions. Using cue-chimeric face stimuli, consisting of half shaded and half SFM-defined surfaces, I have found that classic holistic-face effects such as the composite face effect could also be observed, implying that the brain integrates multiple cues, even if they are spatially segregated, to form a coherent percept which is then subjected to the same errors of perception that normally congruent surfaces are. As such, the perception of one part of a facial surface is affected by the perception of the other part of a facial surface, even if the two parts are composed of different depth cues.

In a related experiment, the perception of an SFM-defined facial surface is obliterated if the SFM-defined faces is overlaid on a shaded face, even if all the dots composing the SFM-defined face are clearly visible. The SFM-defined face only reappears when the contrast and luminance of the shaded face are low.

Object recognition across the two pathways

It seems that the classic distinctions between the dorsal and ventral pathways are limited, and a range of overlap of function exists between the two pathways, particularly concerning 3-D representation of objects from cues. However, the conscious process of vision is an active one, requiring constant motor manipulations to understand the world around us. One never looks at a new object standing perfectly still—we investigate novelty. The investigation accumulates a wealth of 3-D information, not just about the object of interest, but also how the object interacts with its immediate environment. Our brain must keep this information interlinked and successively build a complex representation of an object that may be used to make predictions about the role of this object with others and its perception under previously-unwitnessed environments and conditions. This interplay between action—spatial manipulation and evaluation of the results of those acts—serves as a reminder that whatever functions we attribute distinctly to the different anatomical pathways, we must in the end bring them together to describe a coherent phenomenon. Natural vision involves constant input of 3-D information—it is never 2-D—and the fact that the actions are in turn guided by the evidence one accumulates about an object suggests that normal object recognition requires a constant interaction between the dorsal and ventral streams as we currently envision them.

References

Afraz, S. R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature, 442*(7103), 692-695.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol, 52*(6), 1106-1130.

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *J Neurophysiol, 51*(1), 16-31.

Andersen, R. A., & Bradley, D. C. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Science, 2*(6), 222-228.

Andersen, R. A., & Siegel, R. M. (1990). Motion processing in primate cortex. In *Signal and Sense: Local and Global Order in Perceptual Maps* (pp. 163-184): John Wiley & Sons.

Anderson, K. C., & Siegel, R. M. (2005). Three-dimensional structure-from-motion selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. *Cereb Cortex, 15*(9), 1299-1307.

Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J Neurosci, 11*(1), 168-190.

Bentin, S., Allison, T., Puce, A., Perez, A., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J Cogn Neurosci, 8*, 551-565.

Blanke, O., Brooks, A., Mercier, M., Spinelli, L., Adriani, M., Lavanchy, L., et al. (2007). Distinct mechanisms of form-from-motion perception in human extrastriate cortex. *Neuropsychologia, 45*(4), 644-653.

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex, 8*(6), 510-523.

Borrmann, K., Boutet, I., & Chaudhuri, A. (2003). Spatial attention favors faces over non-face objects in an attentional cueing task. *Journal of Vision, 3*(9), 818.

Britten, K. H., Newsome, W. T., & Saunders, R. C. (1992). Effects of inferotemporal cortex lesions on form-from-motion discrimination in monkeys. *Exp Brain Res, 88*(2), 292-302.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol, 46*(2), 369-384.

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: the role of dynamic information in facial recognition. In M. M. Gruneberg, P. E. Morris & R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues* (pp. 169–174): John Wiley & Sons.

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalizing across different views of faces. *Vision Res, 47*(16), 2164-2169.

Burkhalter, A., & Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J Neurosci, 6*(8), 2327-2351.

Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J Neurophysiol, 71*(6), 2269-2280.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Mem Cognit, 26*(4), 780-790.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol, 75*(3), 1306-1308.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology, 32*(4), 331-341.

De Weerd, P., Desimone, R., & Ungerleider, L. G. (1996). Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Vis Neurosci, 13*(3), 529-538.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci, 4*(8), 2051-2062.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol, 57*(3), 835-868.

Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Res, 25*(3), 441-452.

DeYoe, E. A., & Van Essen, D. C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature, 317*(6032), 58-61.

Dijkerman, H. C., Milner, A. D., & Carey, D. P. (1999). Motion parallax enables depth processing for action in a visual form agnosic when binocular vision is unavailable. *Neuropsychologia, 37*(13), 1505-1510.

Dupont, P., De Bruyn, B., Vandenberghe, R., Rosier, A. M., Michiels, J., Marchal, G., et al. (1997). The kinetic occipital region in human visual cortex. *Cereb Cortex, 7*(3), 283-292.

Durand, J. B., Nelissen, K., Joly, O., Wardak, C., Todd, J. T., Norman, J. F., et al. (2007). Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron, 55*(3), 493-505.

Dursteler, M. R., & Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *J Neurophysiol, 60*(3), 940-965.

Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron, 45*(5), 793-800.

Farivar, R., Germann, J., Petrides, M., Blanke, O., & Chaudhuri, A. (2006). *Dorsoventral integration for recognizing motion-defined faces and objects: evidence from psychophysics, neuropsychology, and functional imaging.* Paper presented at the Society for Neuroscience 2006, Atlanta, GA.

Ferrera, V. P., Rudolph, K. K., & Maunsell, J. H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *J Neurosci, 14*(10), 6171-6186.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360*(6402), 343-346.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol, 76*(4), 2718-2739.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*, 207-208.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci, 3*(2), 191-197.

Gegenfurtner, K. R., Kiper, D. C., & Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *J Neurophysiol, 77*(4), 1906-1923.

Gilaie-Dotan, S., Ullman, S., Kushnir, T., & Malach, R. (2002). Shape-selective stereo processing in human object-related visual areas. *Hum Brain Mapp, 15*(2), 67-79.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci, 15*(1), 20-25.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349*(6305), 154-156.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst), 107*(1-3), 293-321.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166*(910), 1303-1306.

Grunewald, A., Bradley, D. C., & Andersen, R. A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *J Neurosci, 22*(14), 6195-6207.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nat Neurosci, 1*(3), 235-241.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci U S A, 88*(5), 1621-1625.

Hegde, J., & Felleman, D. J. (2007). Reappraising the functional implications of the primate visual anatomical hierarchy. *Neuroscientist, 13*(5), 416-421.

Heywood, C. A., Gadotti, A., & Cowey, A. (1992). Cortical area V4 and its role in the perception of color. *J Neurosci, 12*(10), 4056-4065.

Hilgetag, C. C., O'Neill, M. A., & Young, M. P. (1996). Indeterminate organization of the visual system. *Science, 271*(5250), 776-777.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Curr Biol, 11*(11), 880-885.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*(2), 201-222.

Holmes, G., & Horax, G. (1919). Disturbances of spatial orientation and visual attention with loss of stereoscopic vision. *Archives of Neurology and Psychiatry, 1*, 385-407.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A, 96*(16), 9379-9384.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2002). Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron, 35*(4), 793-801.

Janssen, P., Vogels, R., & Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science, 288*(5473), 2054-2056.

Jiang, F., Blanz, V., & O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation: A view from the other side of the mean. *Psychol Sci, 17*(6), 493-500.

Jiang, F., Blanz, V., & O'Toole, A. J. (2007). The role of familiarity in three-dimensional view-transferability of face identity adaptation. *Vision Res, 47*(4), 525-531.

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain, 93*(4), 793-820.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

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*(1476), 2109-2128.

Klüver, H., & Bucy, P. C. (1937). "Psychic blindness" and other symptoms following bilateral temporal lobectomy in Rhesus monkeys. *American Journal of Physiology, 119*, 352-353.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci, 11*(2), 224-231.

Kourtzi, Z., Bulthoff, H. H., Erb, M., & Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. *Nat Neurosci, 5*(1), 17-18.

Kriegeskorte, N., Sorger, B., Naumer, M., Schwarzbach, J., van den Boogert, E., Hussy, W., et al. (2003). Human cortical object recognition from a visual motion flowfield. *J Neurosci, 23*(4), 1451-1463.

Laeng, B., & Caviness, V. S. (2001). Prosopagnosia as a deficit in encoding curved surface. *J Cogn Neurosci, 13*(5), 556-576.

Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2007). Attention capture by faces. *Cognition*.

Lappe, M., Bremmer, F., Pekel, M., Thiele, A., & Hoffmann, K. P. (1996). Optic flow processing in monkey STS: a theoretical and experimental approach. *J Neurosci, 16*(19), 6265-6285.

Lehky, S. R., & Sereno, A. B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. *J Neurophysiol, 97*(1), 307-319.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci, 4*(1), 89-94.

Liu, C. H., Collin, C. A., Farivar, R., & Chaudhuri, A. (2005). Recognizing faces defined by texture gradients. *Percept Psychophys, 67*(1), 158-167.

Liu, C. H., & Ward, J. (2006). The use of 3D information in face recognition. *Vision Res, 46*(6-7), 768-773.

Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annu Rev Neurosci, 16*, 369-402.

Mishkin, M. (1954). Visual discrimination performance following partial ablations of the temporal lobe. II. Ventral surface vs. hippocampus. *J Comp Physiol Psychol, 47*(3), 187-193.

Mishkin, M., & Pribram, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe. I. Ventral vs. lateral. *J Comp Physiol Psychol, 47*(1), 14-20.

Murasugi, C. M., Salzman, C. D., & Newsome, W. T. (1993). Microstimulation in visual area MT: effects of varying pulse amplitude and frequency. *J Neurosci, 13*(4), 1719-1729.

Murray, S. O., Olshausen, B. A., & Woods, D. L. (2003). Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb Cortex, 13*(5), 508-516.

Nakamura, H., Kuroda, T., Wakita, M., Kusunoki, M., Kato, A., Mikami, A., et al. (2001). From three-dimensional space vision to prehensile hand movements: the lateral intraparietal area links the area V3A and the anterior intraparietal area in macaques. *J Neurosci, 21*(20), 8174-8187.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci, 8*(6), 2201-2211.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *J Cogn Neurosci, 12*(6), 1013-1023.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci, 6*(6), 261-266.

Orban, G. A., Claeys, K., Nelissen, K., Smans, R., Sunaert, S., Todd, J. T., et al. (2005). Mapping the parietal cortex of human and non-human primates. *Neuropsychologia, 44*, 2647-67

Orban, G. A., Janssen, P., & Vogels, R. (2006). Extracting 3D structure from disparity. *Trends Neurosci, 29*(8), 466-473.

Orban, G. A., Sunaert, S., Todd, J. T., Van Hecke, P., & Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. *Neuron, 24*(4), 929-940.

Paradis, A. L., Cornilleau-Peres, V., Droulez, J., Van De Moortele, P. F., Lobel, E., Berthoz, A., et al. (2000). Visual perception of motion and 3-D structure from motion: an fMRI study. *Cereb Cortex, 10*(8), 772-783.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *J Neurophysiol, 82*(5), 2490-2502.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: do we know more now than we did 20 years ago? *Annu Rev Psychol, 58*, 75-96.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res, 47*(3), 329-342.

Peuskens, H., Claeys, K. G., Todd, J. T., Norman, J. F., Van Hecke, P., & Orban, G. A. (2004). Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J Cogn Neurosci, 16*(4), 665-682.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, A. K. C. (1997). Recognizing Moving Faces: The Relative Contribution of Motion and Perspective View Information. *Visual Cognition, 4*, 409-438.

Pohl, W. (1973). Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *J Comp Physiol Psychol, 82*(2), 227-239.

Priftis, K., Rusconi, E., Umilta, C., & Zorzi, M. (2003). Pure agnosia for mirror stimuli after right inferior parietal lesion. *Brain, 126*(Pt 4), 908-919.

Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol, 78*(6), 2904-2915.

Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Curr Opin Neurobiol, 16*(4), 408-414.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nat Neurosci, 3 Suppl*, 1199-1204.

Rizzo, M., & Nawrot, M. (1998). Perception of movement and shape in Alzheimer's disease. *Brain, 121 (Pt 12)*, 2259-2270.

Roe, A. W., & Ts'o, D. Y. (1995). Visual topography in primate V2: multiple representation across functional stripes. *J Neurosci, 15*(5 Pt 2), 3689-3715.

Rolls, E. T. (2004). Invariant Object and Face Recognition. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (Vol. 2, pp. 1165-1178). Cambridge, MA: MIT Press.

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain, 126*(Pt 11), 2381-2395.

Rossion, B., Kung, C. C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proc Natl Acad Sci U S A, 101*(40), 14521-14526.

Ryu, J. J., & Chaudhuri, A. (2006). Representations of familiar and unfamiliar faces as revealed by viewpoint-aftereffects. *Vision Res, 46*(23), 4059-4063.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature, 346*(6280), 174-177.

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. *J Neurosci, 12*(6), 2331-2355.

Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron, 49*(2), 307-318.

Schiller, P. H. (1995). Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature, 376*(6538), 342-344.

Schlack, A., & Albright, T. D. (2007). Remembering visual motion: neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron, 53*(6), 881-890.

Seltzer, B., & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res, 149*(1), 1-24.

Seltzer, B., & Pandya, D. N. (1991). Post-rolandic cortical projections of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol, 312*(4), 625-640.

Shikata, E., Hamzei, F., Glauche, V., Knab, R., Dettmers, C., Weiller, C., et al. (2001). Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. *J Neurophysiol, 85*(3), 1309-1314.

Shikata, E., Hamzei, F., Glauche, V., Koch, M., Weiller, C., Binkofski, F., et al. (2003). Functional properties and interaction of the anterior and posterior intraparietal areas in humans. *Eur J Neurosci, 17*(5), 1105-1110.

Shikata, E., Tanaka, Y., Nakamura, H., Taira, M., & Sakata, H. (1996). Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. *Neuroreport, 7*(14), 2389-2394.

Shipp, S., & Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature, 315*(6017), 322-325.

Shipp, S., & Zeki, S. (1989). The Organization of Connections between Areas V5 and V2 in Macaque Monkey Visual Cortex. *Eur J Neurosci, 1*(4), 333-354.

Siegel, R. M., & Andersen, R. A. (1988). Perception of three-dimensional structure from motion in monkey and man. *Nature, 331*(6153), 259-261.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron, 24*(1), 49-65, 111-125.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *J Neurosci, 11*(9), 2768-2785.

Steeves, J. K., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., et al. (2006). The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia, 44*(4), 594-609.

Sugihara, H., Murakami, I., Shenoy, K. V., Andersen, R. A., & Komatsu, H. (2002). Response of MSTd neurons to simulated 3D orientation of rotating planes. *J Neurophysiol, 87*(1), 273-285.

Tanaka, K., Sugita, Y., Moriya, M., & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *J Neurophysiol, 69*(1), 128-142.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends Cogn Sci, 7*(1), 23-30.

Tolias, A. S., Keliris, G. A., Smirnakis, S. M., & Logothetis, N. K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. *Nat Neurosci, 8*(5), 591-593.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17*, 257-279.

Troje, N. F., & Bulthoff, H. H. (1996). Face recognition under varying poses: the role of texture and shape. *Vision Res, 36*(12), 1761-1771.

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*(4), 483-487.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nat Neurosci, 6*(9), 989-995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*(5761), 670-674.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci, 4*(8), 832-838.

Tsutsui, K., Jiang, M., Sakata, H., & Taira, M. (2003). Short-term memory and perceptual decision for three-dimensional visual features in the caudal intraparietal sulcus (Area CIP). *J Neurosci, 23*(13), 5486-5495.

Tsutsui, K., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J Neurophysiol, 86*(6), 2856-2867.

Tyler, C. W., Likova, L. T., Kontsevich, L. L., & Wade, A. R. (2006). The specificity of cortical region KO to depth structure. *Neuroimage, 30*(1), 228-238.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge: MIT Press.

Vaina, L. M. (1989). Selective impairment of visual motion interpretation following lesions of the right occipito-parietal area in humans. *Biol Cybern, 61*(5), 347-359.

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: a case study. *Vis Neurosci, 5*(4), 353-369.

Valyear, K. F., Culham, J. C., Sharif, N., Westwood, D., & Goodale, M. A. (2006). A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia, 44*(2), 218-228.

Van Oostende, S., Sunaert, S., Van Hecke, P., Marchal, G., & Orban, G. A. (1997). The kinetic occipital (KO) region in man: an fMRI study. *Cereb Cortex, 7*(7), 690-701.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., et al. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science, 298*(5592), 413-415.

Wang, G., Tanifuji, M., & Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res, 32*(1), 33-46.

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb Cortex, 4*(5), 470-483.

Weigelt, S., Kourtzi, Z., Kohler, A., Singer, W., & Muckli, L. (2007). The cortical representation of objects rotating in depth. *J Neurosci, 27*(14), 3864-3874.

Welchman, A. E., Deubelius, A., Conrad, V., Bulthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci, 8*(6), 820-827.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cereb Cortex, 15*(8), 1234-1242.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.

Young, M. P. (1992). Objective analysis of the topological organization of the primate cortical visual system. *Nature, 358*(6382), 152-155.

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

Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain, 106 (Pt 2)*, 313-340.