2015

Effects of Temporal and Spatial Context Within the Macaque Face-Processing System

Clark Andrew Fisher

EFFECTS OF TEMPORAL AND SPATIAL CONTEXT WITHIN THE MACAQUE FACE-PROCESSING SYSTEM

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Clark Andrew Fisher

June 2015

EFFECTS OF TEMPORAL AND SPATIAL CONTEXT WITHIN THE MACAQUE FACE-PROCESSING SYSTEM

Clark Andrew Fisher, Ph.D.

The Rockefeller University 2015

Temporal and spatial context play a key role in vision as a whole, and in face perception specifically. However, little is known about the neurophysiological mechanisms by which contextual cues exert their effects. Anatomically distinct face patches in the macaque brain analyze facial form, and studies of the activity within these patches have begun to clarify the neural machinery that underlies facial perception. This system provides a uniquely valuable opportunity to study how context affects the perception of form. We used functional magnetic resonance imaging (fMRI) to investigate the brain activity of macaque monkeys while they viewed faces placed in either temporal or spatial context.

Facial motion transmits rich and ethologically vital information, but the way that the brain interprets such natural temporal context is poorly understood. Facial motion activates the face patches and surrounding areas, yet it is not known whether this motion is processed by its own specialized neural machinery, and if so, what that machinery's organization might be. To address these questions, we monitored the brain activity of macaque monkeys while they viewed low- and high-level motion and form stimuli. We found that, beyond classical motion areas and the known face patch system, moving faces recruited a heretofore-unrecognized face patch. Although all face patches displayed distinctive selectivity for face motion over object motion, only two face patches

preferred naturally moving faces, while three others preferred randomized, rapidly varying sequences of facial form. This functional divide was anatomically specific, segregating dorsal from ventral face patches, thereby revealing a new organizational principle of the macaque face-processing system.

Like facial motion, bodies can provide valuable social context, revealing emotion and identity. Little is known about the joint processing of faces and bodies, even though there is reason to believe that their neural representations are intertwined. To identify interaction between the neural representations of face and body, we monitored the brain activity of the same monkeys while they viewed pictures of whole monkeys, isolated monkey heads, and isolated monkey bodies. We found that certain areas, including anterior face patches, responded more to whole monkeys than would be predicted by summing the separate responses to isolated heads and isolated bodies. The supralinear response was specific to viewing the conjunction of head and body; heads placed atop non-body objects did not evoke this activity signature. However, a supralinear context response was elicited by pixelated, ambiguous faces presented on bodies. The size of this response suggests that the supralinear signal in this case did not result from the disambiguation of the ambiguous faces.

These studies of contextually evoked activity within the macaque face processing system deepen our understanding of the cortical organization of both visual context and face processing, and identify promising sites for future research into the mechanisms underlying these critical aspects of perception.

Acknowledgements

Although The Rockefeller University's guidelines insist that my name stand alone on cover page of this thesis, I have never before produced work which has been so directly dependent on the efforts and advice of so many others.

I would like to start off by thanking Dr. Winrich Freiwald, my advisor and mentor whose influence is clear in every question I pose and experiment I run. His passion for asking meaningful questions, and for the experimental promise of the macaque face processing system, has inspired me from the first. Just as important to me, though, has been his unwavering commitment to – and engagement with – the success of the students who study in his laboratory.

Many other scientific authorities have lent me vital support along the way. Thanks to my advisory committee – Drs. A. James Hudspeth, Gaby Maimon, and Keith Purpura – for providing a balanced mix of critique and support that has helped me to navigate the inevitable roadblocks and detours of scientific research. Special thanks goes to Dr. Margaret Livingstone for lending her time and expertise as my external examiner. I am also grateful to many scientists whose words have shaped the way I approached my research: Dr. Michael J. Berry, who mentored me as an undergraduate and provided me with the skills, confidence, and desire to continue forward in neuroscience; Dr. Sheila Nirenberg, who welcomed me into her lab and has permanently altered my view of how we can know what the brain knows; Drs. Kalanit Grill-Spector, Douglas Greve, Hauke Kolster, Adam Kohn, Sebastian Moeller, Jonathan Victor, Jonathan Winawer, Lawrence Wald, and Galit Yovel who provided critical technical and conceptual advice throughout my research.

A tremendous thank you to all of the members of the Freiwald lab, a group of scientists that has provided not only verbal, intellectual, and emotional support, but also the physical, logistical, and experimental assistance that was necessary for every single day of research. Thanks to Bryan Baxter, Michael Borisov, Akinori Ebihara, Margaret Fabiszak, Alejandra Gonzalez, Rizwan Huq, Sofia Landi, Dr. Lucy Petro, Pablo Polosecki, Dr. Christina Pressl, Dr. Srivatsun Sadagopan, Dr. Ilaria Sani, Dr. Caspar Schwiedrzik, Stephen Serene, Dr. Stephen Shepherd, Dr. Julia Sliwa, Dr. Sara Steenrod, and Dr. Wilbert Zarco for help with animal training, data collection, and discussion of methods. Thanks also to all of the veterinary services and animal husbandry staff for care of the subjects, staff of the The Rockefeller University shops and production facilities for manufacture of tools and infrastructure, and to everyone at the Citigroup Biomedical Imaging Center for practical support and logistical understanding.

I would not have had the chance to enter a laboratory, nor the strength to remain in one, without the institutional support that has sustained me. I owe my deepest gratitude to Dr. Olaf Andersen, Ruth Gotian, and the rest of the Tri-Institutional MD-PhD Program for providing opportunity, structure, and the sense (which I think can be rare in science and graduate school) that the powers that be have your back. Also, to The Rockefeller University and its Graduate Program for being so uniformly helpful that I am now a bit reluctant to leave.

Thank you to the many generous sources of funding for my research and training, including the NIH (through Medical Scientist Training Program grant NIGMS T32GM007739 to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program and R01 grant EY021594-01A1 to Dr. Freiwald), as well as the Pew Scholars Program, the McKnight Endowment Fund for

Neuroscience, the NSF Science and Technology Center for Brains, Minds and Machines, and the New York Stem Cell Foundation, who have sustained the research in Dr. Freiwald's laboratory.

Thank you Mom and Dad, for being my first and most-enduring role models, for the boundless love and support, and for demonstrating how a person can remain involved in learning, teaching, and research through many different paths. Thank you Miranda, for all of your love and support, sent from as far as Ohio or as close as the other side of the Expedit.

Finally, thank you Marya, for eleven wonderful years of being my companion, co-conspirator, and inspiration, at home and in research. Also: for editing every word of my thesis, including this acknowledgement.

Table of Contents

List of Figures

List of Abbreviations

AD	anterior dorsal (face patch)
AF	anterior fundus (face patch)
AL	anterior lateral (face patch)
AM	anterior medial (face patch)
aMF	anterior middle fundus (face patch)
antSTS	anterior superior temporal sulcus (body patch)
aSTS	anterior superior temporal sulcus (face area)
BOLD	blood-oxygen-level-dependent (imaging)
CBV	cerebral blood volume
CI	confidence interval
EBA	extrastriate body area
FBA	fusiform body area
FDR	false discovery rate
FFA	fusiform face area
fMRI	functional magnetic resonance imaging
fps	frames per second
FST	fundus of superior temporal (visual area)
IT	inferotemporal (cortex)
LST	lower superior temporal (visual area)
MD	middle dorsal (face patch)
MF	middle fundus (face patch)
midSTS	middle superior temporal sulcus (body patch)
ML	middle lateral (face patch)

MRI	magnetic resonance imaging
MST	medial superior temporal (visual area)
MT	middle temporal (visual area)
MVPA	multi-voxel pattern analysis
OFA	occipital face area
PL	posterior lateral (face patch)
pPL	posterior posterior lateral (face patch)
pSTS	posterior superior temporal sulcus (face area)
ROI	region of interest
STS	superior temporal sulcus
VASO	vascular-space-occupancy (imaging)

1. Introduction

Making sense of the rapidly changing visual world is a daunting challenge, and one of the brain's great feats. The brain works hard to interpret the spots of light and dark, contrast and color captured by the retina, and to ultimately assign them to objects as diverse as a candle's flame, a missing set of keys, and the smiling face of a friend. One way that the brain can make this thorny problem more manageable is to exploit temporal and spatial contextual cues. For instance, the dancing, flickering motion of flame is shared by few other forms, and a novelty keychain can be a strong signal that your keys are hiding somewhere nearby.

The wide gulf that exists between basic visual percepts and a meaningful visual world may be most evident when the normal functions of the brain break down. It is in these circumstances that the crucial role played by context is clearest. Consider the case of Dr. P, the music teacher who inspired the title of neurologist Oliver Sacks' popular book, *The Man Who Mistook His Wife for a Hat* (Sacks 1998). When Sacks flipped through a *National Geographic* and asked Dr. P to describe the pictures,

> …a striking brightness, a colour, a shape would arrest his attention and elicit comment—but in no case did he get the scene-as-a-whole. He failed to see the whole, seeing only details, which he spotted like blips on a radar screen. He never entered into relation with the picture as a whole—never faced, so to speak, its physiognomy. He had no sense whatever of a landscape or scene.

This inability to create a meaningful visual whole from its constituent parts had its most dramatic effect on the perception of faces, where it meant that he could recognize no one by sight, "neither his family, nor his colleagues, nor his pupils, nor himself."

To mitigate the major social deficit caused by his form-blindness, Dr. P could still occasionally identify those around him by leveraging visual context. Sometimes, he relied upon the temporal context provided by motion; Sacks explains that "…though he could not recognise his students if they sat still, if they were merely 'images', he might suddenly recognise them if they *moved*. 'That's Karl,' he would cry. 'I know his movements, his body-music'" (Sacks 1998). At other times, Dr. P relied upon the spatial context provided by a characteristic feature near – or even on – the face: "Churchill's cigar, Schnozzle's nose: as soon as he had picked out a key feature he could identify the face."

While the brain relies on contextual clues like these to generate the sense of vision, knowing this does not explain *how* the neural underpinnings of vision function. In fact, the realization that contextual information is important presents a new set of challenges for understanding the neural basis of sight. How, for instance, does the brain represent the distinctive movement of Dr. P's student Karl, and how is this dynamic information linked to knowledge of Karl's form, which stays constant even as Karl moves? Does viewing a cigar enhance the activity of neurons that represent the form of Winston Churchill's face, or do all associative links between Churchill and his Cuban exist outside of the brain's visual processing machinery?

To answer these questions, it would be helpful to identify localized cortical regions that underlie such contextual effects. Therefore, rather than studying

cases of visual context in general, it is logical to start by studying the effects of context on the representation of one particular class of object: faces. Using faces to study general aspects of higher visual function is appealing not just because faces themselves are important, but also because the neural machinery of face processing has proven to be uniquely identifiable and accessible. Because of this, faces have repeatedly been objects of critical importance in efforts to understand how the brain makes sense of the visual world. Just as faces were fundamental to Oliver Sacks' understanding the depth of Dr. P's deficits, faces may provide the key to understanding the neural circuitry underlying visual context.

In the early 1970s, the discovery of neurons that responded specifically to images of faces was one of the first clues to the high degree of form selectivity maintained by certain visual neurons, and to the level of configural complexity that could be represented by a single neuron (Gross, Rocha-Miranda, & Bender 1972; Desimone et al. 1984). With the advent of functional neuroimaging in the late 20th century, the field of face processing again found itself at the forefront of visual neuroscience; the identification of a macroscopic brain region that showed a specific sensitivity to face form provided compelling evidence that the visual processing of certain classes of objects depends upon specific, anatomically distinct brain regions (Kanwisher, McDermott, & Chun 1997). More recently, the description of the "face patch" system in macaques (Tsao et al. 2003) has built upon these earlier discoveries, revealing that face-selective regions of cortex represent discrete but tightly interconnected functional areas filled with face-selective neurons (Tsao et al. 2006; Moeller, Freiwald, & Tsao 2008; Freiwald & Tsao 2010). As anatomically similar networks have been discovered with selectivities for body form (Pinsk et al. 2005), learned symbols (Srihasam,

Vincent, & Livingstone 2014), and even qualities like color and curvature (Lafer-Sousa & Conway 2013; Yue et al. 2014), it is becoming increasingly clear that learning how the brain treats faces not only elucidates the processing of these socially critical objects, it also establishes a template for understanding how the brain represents a wide diversity of high-level visual features (Orban, Zhu, & Vanduffel 2014).

While these discoveries have fostered much research into the representation of faces and other objects within the brain, comparatively little is understood about how contextual effects fit into this process. This may be a consequence of the face-processing system's greatest experimental benefit: its striking form-selectivity. A fundamental feature of face processing research has been the repeated finding that a still picture of a face, presented on its own, can drive regions of visual cortex that show little or no response to viewing any other object or scene. This has been a boon to scientists, who can work within the rich but restricted space of face form to understand the rules that underlie neural activity in these regions.

However, when we interact with faces, they are rarely the static, disembodied heads most frequently used by face processing researchers. Instead, faces are dynamic objects, capable of switching attention and affect – and therefore significance – in an instant. Face motion itself can contain information normally thought to derive from the static features of a face, such as the identity or mood of a speaker. Faces are not usually seen on their own, but atop bodies that provide important spatial context; if you see a body, odds are good that there is a face nearby. Bodies also provide important social context, as an intent stare means one thing when it is paired with an outstretched hand, and another

when it is paired with a raised fist. In short, faces are *constantly* surrounded by informative temporal and spatial context.

Given the clear importance of context to visual processing on the one hand, and the preexisting knowledge and experimental accessibility of the macaque face patch system on the other, we decided to explore the impact of temporal and spatial context on face processing in the macaque temporal lobe. This approach held the promise of enriching our comprehension of the neural substrates of both visual context effects and face processing, thereby bringing us closer to an understanding of how the flood of information taken in by the retina is endowed with visual meaning by the brain.

Face-Processing in the Temporal Lobes of Macaques and Humans

Although the recognition of the macaque face patch system is still less than fifteen years old, the particular structure of this system and its ability to bridge the fields of human neuroimaging and monkey electrophysiology have made it a uniquely valuable resource for understanding high level aspects of sight.

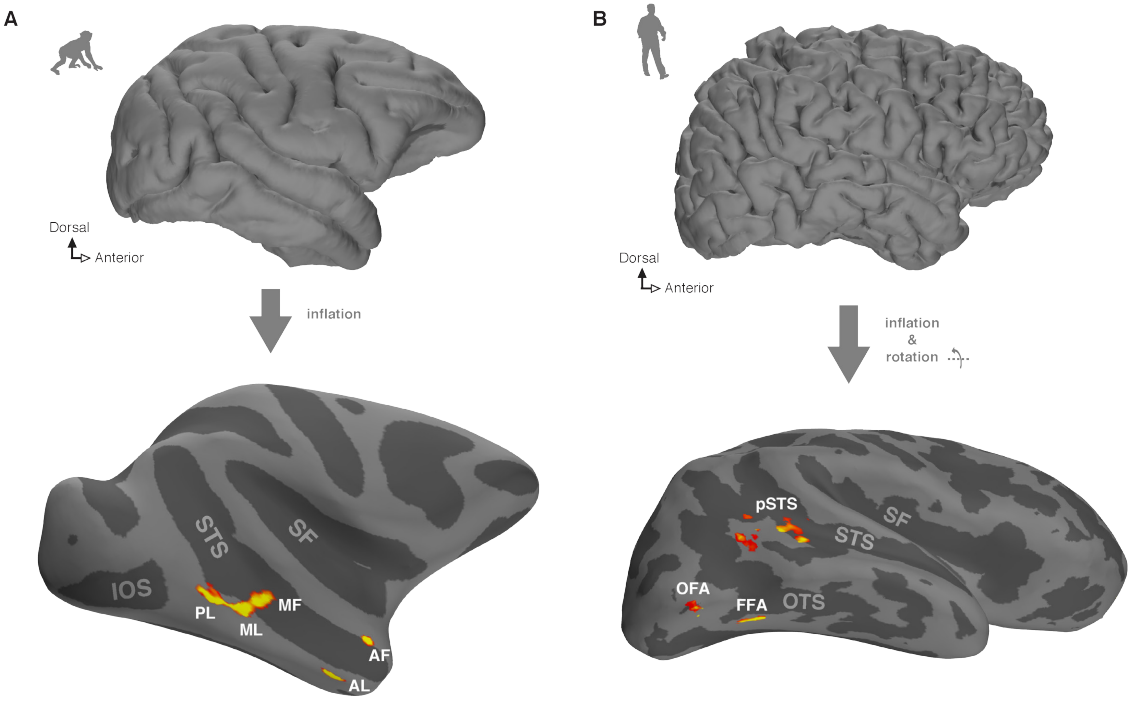
The macaque face patches are regions of cortex, initially identified using functional magnetic resonance imaging (fMRI), that respond more to images of faces than to images of other objects (Tsao et al. 2003). While parcellation and naming conventions have not fully settled, 6 patches along the temporal lobe are increasingly recognized as reproducible, individually identifiable regions (Figure 1A; Tsao, Moeller, & Freiwald 2008; Vanduffel, Zhu, & Orban 2014; Russ & Leopold 2015). Three of these patches sit along the ventrolateral lip of the superior temporal sulcus (STS), running from its middle to its anterior end: the posterior lateral face patch (PL), the middle lateral face patch (ML), and the

Figure 1. Temporal Lobe Face Areas in the Macaque Monkey and Human

The locations of the most commonly reported face-selective areas on models (not to scale) of representative macaque and human brains. These models have been inflated so that areas hidden within sulci can be seen. Dark gray areas represent the deepest points within sulci. Locations of face-selective cortex are marked with red and yellow.

(**A**) Commonly reported face patches along the STS of the macaque temporal lobe. Face patches PL, ML, and AL are found along the ventral lip of the STS. MF and AF are found in the fundus and dorsal bank of the STS. Face patch AM, which is also widely reported, is not visible in this view. It is on the ventral surface of the temporal lobe, anterior and ventral to face patch AL. Grey labels indicate landmark sulci: IOS, inferior occipital sulcus; SF, sylvian fissure; STS, superior temporal sulcus.

(B) Commonly reported face areas in the human brain. The view of the inflated human brain has been rotated 45º in the indicated direction, so that face areas on the ventral surface of the brain are visible. The OFA and FFA are on the ventral surface of the brain. The pSTS face area is within the superior temporal sulcus, and is known to be best-localized using dynamic stimuli. Labels again indicate landmark sulci: OTS, occipitotemporal sulcus; SF, sylvian fissure; STS, superior temporal sulcus.


A
Dorsal
Anterior
inflation
STS
SF
IOS
PL
ML
MF
AF
AL
B
Dorsal
Anterior
inflation
&
rotation
pSTS
SF
STS
OFA
FFA
OTS

anterior lateral face patch (AL). Two additional patches lie in the fundus and upper bank of the STS: the middle fundus face patch (MF) and the anterior fundus face patch (AF). The sixth, anterior medial, face patch (AM) is found on the ventral surface of the anterior temporal lobe near the anterior middle temporal sulcus. Microelectrode stimulation during fMRI (Moeller, Freiwald, & Tsao 2008) and recent anatomical studies (Grimaldi, Saleem, & Tsao 2012; Grimaldi, Saleem, & Tsao 2013) have revealed that the face patches are interconnected with each other. Appreciating this system from the macroscopic level, it is notable as a network of discrete but connected regions that respond specifically to a known and circumscribed set of complex visual forms.

Additional temporal lobe face patches have been reported (AD in Pinsk et al. 2009; vV4 and TF in Ku et al. 2011; pPL and aMF in Janssens et al. 2014), but have not been reliably replicated within and across experiments. Beyond the temporal lobe, regions of face-selectivity have also been found in prefrontal cortex (Tsao et al. 2008; Hadj-Bouziane et al. 2008; Janssens et al. 2014). Relatively little is known about these areas at the moment, and, given their location distant from recognized "visual" cortex, they may not be necessary for basic visual perception (though they may be involved in the processing of social signals, see Hornak, Rolls, & Wade 1996; Watson & Platt 2012; P. Johnston et al. 2013). We will put aside these face areas, and hereafter when we refer to face patches, we mean the reliably identified temporal lobe face patches

At the single neuron level, the organization of face patches makes them an ideal site for the study of visual responses in single cells and small networks. 75%–97% of cells recorded in the face patches are face-selective (varying with patch and selectivity criterion), responding more to images of faces than to

images of other objects (Tsao et al. 2006; Freiwald & Tsao 2010; Issa & DiCarlo 2012; Aparicio 2014). This has proven to be exceedingly useful for the characterization of face-selective cells, as it allows researchers to target electrophysiological recording to areas highly enriched in face-selectivity (e.g. Freiwald, Tsao, & Livingstone 2009; Ohayon, Freiwald, & Tsao 2012; Issa & DiCarlo 2012). However, while neurons across all face patches are face-selective, they are not functionally uniform. Freiwald and colleagues (2010) found that when moving from posterior to anterior, from ML to AL to AM, the neurons that they recorded became increasingly tuned to the identity of the face and increasingly tolerant to the specific angle from which the face was viewed. Thus, different face patches are not merely anatomically distinct, but are functionally distinct as well. Recent research further suggests that neurons in these patches may be fundamentally different than the face-selective neurons outside of them (Taubert et al. 2015). This organization makes it possible to explore the varied forms that facial information takes as it moves through the face processing machinery of the temporal lobe.

Information gleaned from the macaque face patch network has direct implications for the mechanisms of human vision, as there is an organizationally similar, possibly homologous, face processing system along the human temporal lobe. Face areas in humans were established in much the same way that macaque face patches would later be identified: functional imaging was used to recognize areas that preferentially responded to face images. Three face areas have been repeatedly found in human studies (Figure 1B): the fusiform face area (FFA) on the fusiform gyrus of the ventral temporal lobe (Kanwisher, McDermott, & Chun 1997), the occipital face area (OFA) in occipito-temporal cortex (Gauthier et al.

2000), and the posterior STS face area (pSTS, or sometimes STS-FA) in the superior temporal sulcus (Puce et al. 1998; Hoffman & Haxby 2000). It is becoming increasingly clear that there are further functional areas, identifiable either by subdividing these regions (Weiner & Grill-Spector 2010) or examining other parts of the temporal lobe (Pinsk et al. 2009; Rajimehr, J. C. Young, & Tootell 2009), so it remains possible that there is an exact human homologue for each macaque face patch. However, despite repeated attempts to use matched experimental design to homologize the macaque and human face-processing systems (Tsao, Moeller, & Freiwald 2008; Pinsk et al. 2009; Polosecki et al. 2013), the ability to link specific nodes between the two systems has remained elusive (Yovel & Freiwald 2013). Even without an established area-to-area homology, monkey studies draw inspiration from human face-processing research – where it is easier to run basic imaging experiments, train subjects to follow instructions, and receive perceptual reports – and human studies benefit from the insights provided by monkey experiments – which offer a greater range of techniques, including targeted single neuron recordings (Freiwald, Tsao, & Livingstone 2009; Freiwald & Tsao 2010; Issa & DiCarlo 2012), anatomical tracing (Grimaldi, Saleem, & Tsao 2012), and optogenetics (Gerits & Vanduffel 2013; Afraz, Boyden, & DiCarlo 2014).

Research into the macaque face patches also relies upon and informs a diversity of other form-processing research. The discovery of form-selective neurons in temporal cortex predates the description of the face patches by 3 decades (Gross, Rocha-Miranda, & Bender 1972). Since that time, studies of how faces (reviewed in Barraclough & Perrett 2011) and other objects (reviewed in DiCarlo, Zoccolan, & Rust 2012) are represented in temporal cortex at large have

revealed remarkable abilities of individual neurons (e.g. Perrett et al. 1989; Oram & Perrett 1996; Booth & Rolls 1998) and shed light on the general organization of the cortical region that is home to the face patches (Connor, Brincat, & Pasupathy 2007; Rust & DiCarlo 2010).

However, it is difficult to characterize the tuning of a form-selective neuron without any intuition of the form to which it responds or the visual dimensions along which it is tuned, limiting the questions that can be asked. While the study of face-selective neurons offers one answer to this challenge (as faces have a stereotyped form and a number of natural dimensions to explore), such neurons are difficult to study outside of the face patches, where only 9–35% of neurons are face-selective (Perrett, Rolls, & Caan 1982; Bell et al. 2011; Aparicio 2014). Furthermore, while the organization of earlier visual areas is generally established at this point, the functional organization of the temporal lobe is still a matter of active discussion (Van Essen 2004); even when a neuron, face-selective or otherwise, is characterized, it can be difficult to know how to define a functionally related neural population or know where it fits in a wider processing hierarchy. The face patches not only provide a high density of face-selective neurons and their attendant benefits, they also allow for the targeted study of separate and identifiable functional areas within the poorly delineated landscape of the temporal lobe. Thus, this system supports a more deliberate exploration of how form is processed at the highest levels of the visual system.

Given the benefits of the face patch system, both those inherent to its form and those resulting from decades of human and macaque research, it is a logical place to study the neural processing of visual form. Beyond these general benefits, however, the known perceptual effects of facial motion and body

context make faces the ideal subjects with which to explore how temporal and spatial context help the brain to make sense of the world.

How the Brain Views Moving Faces

Motion can reveal a great deal about an object's form, state, and identity. At a very rudimentary level, this happens when an object's motion exposes previously hidden views of its shape. But motion also has the ability to embody dynamic signatures and properties (for instance speed, flexibility, and animacy) that can only be suggested by a single, frozen pose. It therefore makes intuitive sense that the brain would make use of temporal context to understand visual forms. In the case of faces, this intuition has been verified, revealing that the brain relies upon facial motion – and not just facial form – to gather critical social information.

Emotion, for instance, can be inferred from facial movements (reviewed in Krumhuber, Kappas, & Manstead 2013). This is especially true in cases where facial form is degraded or absent. For example, when a face is covered with white spots and all other form cues are removed, emotions can be clearly identified from the moving dots (Bassili 1978; Bassili 1979). This is true even though it is difficult to identify a face from any still image of the dots. While motion's benefit is less evident when the full form of a face is available (e.g. Cunningham & Wallraven 2009; Fiorentini & Viviani 2011), it is still relied upon in certain instances, as when emotions are subtle (Ambadar, Schooler, & Cohn 2005) or easily confused (Lemay, Kirouac, & Lacouture 1995; Wehrle et al. 2000). Facial movements can also convey identity information (reviewed in Xiao et al. 2014). This is perhaps more surprising, as a person's identity (unlike his

temperament) is fixed, no matter the range motions expressed by his face. Many manipulations that impair the identification of static faces (such as contrast inversion, image inversion, or posterization) still allow the identification of familiar faces in motion (B. Knight & A. Johnston 1997; Lander, Christie, & Bruce 1999). People can even be identified when their motions are used to animate an anonymous computer-generated model (Hill & A. Johnston 2001).

How are the dynamic features that support such recognition encoded in the brain? It seems increasingly likely that they are carried by a pathway functionally distinct from the one encoding static face form. Patients who demonstrate facial processing deficits that are specific to static stimuli are a clear example of this. For instance, there are multiple reports of patients with prosopagnosia (face blindness, as Sacks' Dr. P demonstrated) who can identify specific dynamic faces although they distinguish static faces poorly (Steede, Tree, & Hole 2007; Longmore & Tree 2013; Richoz et al. 2015). Other reports describe people who, after traumatic brain injury, have increased difficulty judging emotions from static faces but not from moving ones (Humphreys, Donnelly, & Riddoch 1993; Adolphs, Tranel, & Damasio 2003; McDonald & Saunders 2005).

Such clinical findings align with popular "division of labor" models of face processing, which suggest that different aspects of faces are processed by separate functional streams within the brain (Bruce & A. W. Young 1986; Haxby, Hoffman, & Gobbini 2000; Weiner & Grill-Spector 2013). These functionally parcellated models of face-processing echo the dominant understanding of the wider visual system, which holds that high-level visual processing is split into at least two primary streams: a ventral stream along the occipital and temporal lobes that analyzes object form – the "what" pathway – and a dorsal stream

leading into the parietal lobe that analyzes object location and movements – the "where" or "how" pathway (Mishkin, Ungerleider, & Macko 1983; Goodale & Milner 1992).

In the human brain, functional imaging has begun to identify the anatomical location of "face motion" areas that could represent a functional stream for the processing of dynamic facial features. Puce and colleagues (1998) were the first to suggest that a region in the posterior STS may be involved in processing facial motion. They found that this region specifically responded to eye and mouth movements, but not to the motion of geometrical patterns. This region, now sometimes called the pSTS face area (see discussion above), is also particularly activated when attention is directed to the direction of gaze, a changeable aspect of the face (Hoffman & Haxby 2000). In agreement with division of labor models, pSTS's sensitivity to dynamic features does not appear to be a general property of the face processing system. While pSTS and a more-recently discovered anterior STS face area (aSTS) are specifically engaged by facial motion, more ventral face-selective areas (including OFA and FFA) have been found to respond almost equally to videos of and static pictures of faces (Pitcher et al. 2011; Polosecki et al. 2013; but see Fox, Iaria, & Barton 2009; Schultz & Pilz 2009; Schultz et al. 2013). A recent study suggests that the naturalness of face motion is an important driver of the dorsal areas' response to motion, as scrambling the order of a face movie's frames decreases the resulting activity in the face-responsive STS, but not in the OFA or FFA (Schultz et al. 2013). Therefore, the emerging picture of facial motion processing in the human brain centers around dorsal areas that are activated by the natural dynamics of faces;

these areas stand in contrast to ventral face areas, such as the OFA and FFA, which are selective for face form but largely unaffected by facial motion.

While human fMRI experiments have provided a provisional indication of the areas where face motion processing occurs, it is difficult to gain an understanding of the neural mechanisms at work within these areas. fMRI does not provide the temporal or spatial resolution necessary to probe single-cell or small network activity, and neither targeted electrophysiology nor causal studies are feasible in humans. The macaque monkey provides an appealing experimental alternative, as the localization of facial motion areas in these animals could be subsequently augmented with a variety of fine-scale studies.

In fact, single-neuron electrophysiology has already been used to explore responses to moving biological forms within the macaque brain, though never within an identified face patch. Much of this research has focused on the representation of moving bodies rather than moving faces. For instance, one recent study found that neurons in the STS and neighboring cortex carry enough information to encode both the kinematics of an action and the form of the body performing the action (Singer & Sheinberg 2010). Another found two subpopulations of STS neurons that responded to movies of a walking body: a primarily ventral population that was tuned to the individual poses that make up walking, and a primarily dorsal one that was tuned to walking kinematics, responding to movies of forward walking much more than to movies of backward walking (Vangeneugden et al. 2011). These studies highlight the promise of neuron-level studies of object motion representations, but our limited understanding of the heterogeneous selectivity of body-selective neurons

(Popivanov et al. 2014) makes it difficult to examine the range of responses to body form and motion in a systematic way.

Only two experiments that we know of have explored single neuron responses to face motion, and both did so in a limited way. The first showed that certain face-selective cells in the dorsal STS respond to head translation in a direction selective manner (Oram & Perrett 1996). The second showed that some neurons in the same area respond both to specific articulated movements of a head relative to a body and also to static poses that suggest that these motions (Jellema & Perrett 2003). Both of these studies predated the description of the face patches, and barely begin to explain how neural activity represents facial motion.

Identification of regional specializations for facial motion within the macaque face patch system – matching the specializations of the dorsal human face areas – would facilitate a more thorough and systematic study how the macaque temporal lobe extracts meaning from face movements. A number of studies have recently used fMRI to explore the organization of facial motion processing in the macaque brain. However, none of these have identified areas specialized for natural facial motion like the ones found in humans.

Furl and colleagues (2012) were the first group to apply macaque fMRI to this problem, identifying "motion in face" areas that responded more to videos of facial expressions than to still faces. However, functional areas known to process motion with little regard for underlying form, including areas MT, MST, and FST, are found in the STS fundus neighboring the face patches (Vanduffel et al. 2001; Nelissen, Vanduffel, & Orban 2006). Because moving faces presumably recruit form-agnostic areas such as these, the motion-in-face areas were not defined in a category-specific manner. While the authors of this study were

aware of this issue (stating that they "expected [their contrast] would reliably elicit motion areas known to be relatively domain-general"), this experimental design precludes using this study to identify areas selective for *face* motion.

A subsequent study from our lab addressed this issue, and used fMRI to query the responses of face patches to moving and still faces and to moving and still objects (Polosecki et al. 2013). This study found that all macaque face patches responded more to moving faces than to static ones. However, these patches also tended to respond more to moving non-face objects than to static non-face objects. The authors concluded that most face patches did not have special selectivity for face motion, but instead demonstrated a general motion sensitivity resulting from an overlap with general motion areas in the STS. While the authors did identify two ventral patches (AL and AM) that demonstrated statistically significant face-specific motion responses, the small size of these effects raised the possibility the difference between these patches and all others was just a matter of experimental power.

Another recent fMRI study has approached the problem of the macaque's facial motion response in a somewhat different way. Russ and Leopold (2015) scanned macaques while showing them naturalistic videos that included other macaques interacting with their environments. The researchers scored each moment of each video for different features (including motion energy and the presence of faces), and used the resulting time-courses as linear regressors to explain the recorded fMRI activity. They found that their motion energy regressor explained a great deal of activity within all of the face patches, often more than the face regressor did. However, it is important to note that this was only true when the observed videos included faces. When videos without

animals were used as stimuli, the face patches did not respond. This raises the question of whether the precedence of motion signals in the face patch system resulted from motion per se. Perhaps it resulted, instead, from the diverse views of faces that accompanied motion, a feature that was not captured by the face regressor. In any case, this experiment revealed no evident sub-specialization within the face patches for natural motion.

The repeated inability to identify specific macaque face patches specialized for facial motion throws into doubt not only the utility of the macaque face processing system as an informative homologue of the human one, but also whether the processing of face dynamics in macaques is as experimentally accessible as the processing of face form is. However, human studies suggest two strategies that could allow us to unmask any motion specializations that do exist within the macaque face processing system. First, it is often necessary to use dynamic stimuli to localize motion-selective areas. This common-sense idea is supported by human studies that have identified certain face areas – especially motion-selective ones – only when dynamic stimuli are used to look for them (Fox, Iaria, & Barton 2009; Pitcher et al. 2011). As previous macaque fMRI studies have relied upon static stimuli to identify face patches, they may have missed regions that are particularly selective for motion. Human studies also suggest that it can be difficult to separate the easily entangled variables of natural motion and static information. Because movies contain more static frames than still images, face areas that are widely regarded as form-specific have sometimes been found to be more active in response to movies than in response to still faces (Fox, Iaria, & Barton 2009; Schultz & Pilz 2009; Schultz et al. 2013). However, by comparing the activity elicited by natural face movies to the activity elicited by

temporally scrambled versions of these movies, the distinct sensitivities to dynamic and static information within different face areas can be revealed (Schultz et al. 2013).

By following these leads, it may now be possible to identify areas specialized for processing face motion within the macaque monkey brain. This could help to explain how the brain can extract important social meaning from facial motion, and ultimately how temporal context allows us to fully experience the visual world.

The Role of Body Context in Face Processing

Spatial context, like motion, is a powerful tool for understanding ambiguous and confusing visual information. The brain takes advantage of this, and therefore the specific object that a person sees, and how easily that person sees it, can depend upon other objects within the surrounding scene (reviewed in Bar 2004). Consider a metallic cylinder adorned with a bright logo. If this cylinder is the size of a person's hand and a woman is bringing it to her lips, it is probably a beverage can. If, on the other hand, it is embedded in a remote control's back, it is much more likely to be a battery. As it turns out, bodies provide this type of context for faces and can subconsciously influence what we think about them.

Put in practical terms, information that we think we gather from the form of a face sometimes derives from the form of an associated body instead. Recent research shows that body context exerts a particularly strong influence on the interpretation of emotion, a feature that has long been associated with facial expressions (reviewed in Hassin, Aviezer, & Bentin 2013). The first scientific evidence of this came from studies showing that the emotion perceived from a

face can be modulated by the posture of an associated body (Meeren, van Heijnsbergen, & de Gelder 2005; Aviezer et al. 2008). Interestingly, this effect occurs even when the person judging the face is informed of this modulation and specifically instructed to *ignore* the body (Aviezer et al. 2011). A recent and dramatic example of this effect comes from Aviezer et al. (2012), who found that in certain cases of extreme emotion, still pictures of facial expression provide *no* information for distinguishing a positive emotion from a negative one. They showed that bodies, on the other hand, can be strong indicators of the difference in valence. Despite this, many people judging the emotions of ambiguous emotional faces paired with bodies reported relying on facial cues to categorize emotions, even though their performance showed their choices were dependent on the bodies. The authors refer to this phenomenon as "illusory facial affect", the belief that a face contained emotional information that was actually coming from its body.

Although it is more challenging to demonstrate, body context can also play a similar role in identity recognition. To prove this, Rice and colleagues (Rice, Phillips, & O'Toole 2013) used a machine vision algorithm to identify 100 pairs of face pictures where the faces provided very little information as to whether the pair represented the same person or two different people. While experimental subjects were (unsurprisingly) bad at determining whether each pair showed the same person when faces were presented alone, they did much better when bodies were provided. Again, however, they often reported that they relied upon face features to make identification decisions even though their responses showed otherwise. These experiments into the perception of emotion and

identity not only show that bodies can convey social information, but reveal that this body information is regularly misattributed to faces.

These misattributions could be purely cognitive, the result of an assumption that faces – not bodies – convey emotion and identity. Alternatively, they could reflect the fact that viewing bodies fundamentally alters the neural representations of faces. While the second possibility may seem less likely, psychology experiments have lent it support by demonstrating how bodies can impact face-processing mechanisms. One study found that exposure to a body can change the perception of subsequently viewed face (Ghuman, McDaniel, & Martin 2010). For example, adaptation to pictures of female bodies made a face appear more male, while adaptation to pictures of male bodies did the opposite. Another study showed that bodies paired with faceless heads are much harder to identify when they are upside-down – a dramatic "inversion effect" that is generally indicative of face processing (Brandman & Yovel 2012). Furthermore, when the researchers presented these faceless bodies briefly and quickly obscured them with a blank screen, viewers generally believed that they had seen a face. These findings suggest not merely that bodies carry some of the same information as faces, but also that bodies can directly engage the machinery of face perception.

This impact that bodies have on face processing stands in conflict with the accepted neuroanatomy of biological form perception, which holds that faces and bodies are processed by neighboring but separate networks within the temporal lobe. Just as fMRI was used to discover face areas, it has also been used to identify areas in the human brain that respond more to pictures of bodies and body parts than to other objects. The most commonly reported body areas in

humans are the extrastriate body area (EBA) at the posterior end of the temporal lobe (Downing et al. 2001) and the fusiform body area (FBA) on the fusiform gyrus (Peelen & Downing 2005; Schwarzlose, Baker, & Kanwisher 2005). Much as with face areas, recent studies suggested that additional body areas beyond these may exist (Weiner & Grill-Spector 2010).

Research paints a similar picture of the macaque brain, with patches of body selectivity generally found neighboring the face patches (Tsao et al. 2003; Pinsk et al. 2005). Two of the larger patches have been given names: the one near the middle face patches (ML and MF) has been labeled midSTS and the one near the anterior face patches (AL and AF) is called antSTS (Popivanov et al. 2012). The neurons within these patches are most often selective for body parts, and those that are not evince high selectivity for other forms (Bell et al. 2011; Popivanov et al. 2014). While some face cells are found within these body patches, the category selectivity of the individual neurons in the face and body patches and the distinct activity peaks of face and body areas in fMRI studies of humans (Schwarzlose, Baker, & Kanwisher 2005; Weiner & Grill-Spector 2011) and macaques (Pinsk et al. 2009; Popivanov et al. 2012) have reinforced the idea that these two categories are processed by two *distinct* systems. Even so, the behavioral results described above go against this framework.

One possible reason for this discrepancy is that the experiments which have identified face and body processing areas – and those that have characterized the neurons within them – tend to present faces independent of bodies and bodies independent of faces. If the influence of bodies upon face processing is contextual in nature, it might only be detectable in situations where at least the suggestion of a face is present with the body. Cox et al. (2004) were the first to

find evidence of such a body context effect with fMRI. They showed subjects both pictures of faces and pictures of blurred faces, where all internal details were removed, while measuring the response in the FFA. The blurred, ambiguous faces elicited less activity than the unambiguous faces. However, when the blurred faces were shown with a body below them, they elicited at least as much activity as the lone unambiguous faces did. The authors reached the conclusion that these "contextually defined faces" activated the FFA much like intrinsically defined faces do, hypothesizing a physiological similar in spirit to the illusory facial affect described by Aviezer and colleagues (2012). Notably, Cox et al. claimed that unambiguous faces did not elicit a greater response from FFA when paired with a body, a claim supported by an earlier study (Kanwisher, Stanley, & Harris 1999). Thus, these authors posited that bodies activate face processing by disambiguating an unclear face, and will neither activate face processing when little possibility of a face exists nor when a clear face is visible.

Since this experiment, a number of groups have revisited the question of whether body context affects cortical face responses even when facial form is clear and evident. These groups have taken four fundamentally different approaches to this question. While their conclusions have varied, many of these studies suggest that face responses can be modulated by the presence of a body.

Two studies took the straightforward approach to this question: showing faces and bodies, together and separately, and comparing the evoked activity in cortical regions of interest (Morris, Pelphrey, & McCarthy 2006; Song et al. 2013). Both of these studies demonstrated that showing a body along with a face could increase the activity within the FFA. The one study that examined the OFA (Song et al. 2013), however, found the addition of a body to a face did not significantly

increase OFA activity. In no case did the level of activity elicited by a face and a body rise above the sum of the responses to faces and bodies presented independently. Therefore, while these experiments suggested a response to faces with bodies in FFA, it is possible that the observed activity was merely the sum of independent responses to the two forms, perhaps resulting from overlapping face and body areas.

This led one of the groups (Song et al. 2013) to apply the second approach to this question: multi-voxel pattern analysis (MVPA). The researchers considered whether the spatial pattern of activity within the face areas was fundamentally altered when a body was shown with a face, or if this pattern could be accurately modeled by summing the patterns evoked by faces and bodies presented alone. They concluded that the whole person pattern, at least in the right FFA, was not well explained by the sum of the face and body patterns, and thus must represent a functional integration of body and face information in this area. However, a subsequent study taking a similar MVPA approach found that the response pattern in the fusiform gyrus evoked by pictures of whole people *could* be accurately modeled by a summation of face and body patterns (Kaiser et al. 2014).

Two other studies took a third approach to determine if face and body form are jointly represented in the brain, utilizing fMRI adaptation experiments. Such adaptation experiments are based on the idea that repeated presentations of a stimulus perceived as identical will lead to diminished fMRI responses (which is to say adaptation; Grill-Spector & Malach 2001). Therefore, if a stimulus change releases adaptation within a certain brain area, that brain area must be tuned to whatever feature was changed. One study examined whether changing the body

presented with a face released face areas from adaptation, and found that this occurred in the FFA but not the OFA (Andrews et al. 2010). The second fMRI adaptation experiment used a voxel-wise analysis, and found that some voxels, both near the OFA and near the FFA, were released from adaptation if either the face or the body was changed (Schmalzl, Zopf, & Williams 2012). Furthermore, some even showed supralinear adaptation when the face and body were constant, an adaptation effect that was greater than the one predicted by summing the adaptation effects of holding just the face or just the body constant. Thus, both adaptation experiments suggest a joint response to faces and bodies around face areas.

The most recent experiment to address the interaction between face representations and body context considered whether bodies and faces compete with each other for neural resources (Bernstein et al. 2014). This study built upon experimental results suggesting that distinct objects presented simultaneously will suppress each others' fMRI responses, while objects that the brain groups together show no such competition (McMains & Kastner 2010). The researchers found that body images competed with paired head images in the OFA but not in the FFA, and concluded that head and body are represented holistically in this more anterior face patch.

Some of these studies also sought to determine if faces could exert a contextual effect upon body representations, and examined whether body areas showed evidence of face effects (Morris, Pelphrey, & McCarthy 2006; Song et al. 2013; Bernstein et al. 2014). These experiments generally found no effect of adding faces to bodies in the more posterior EBA. However, the studies that examined the more anterior FBA found evidence of joint tuning, similar to what

the same studies found in the face-selective FFA (Song et al. 2013; Bernstein et al. 2014).

Although their methods and conclusions varied, these studies generally suggest that bodies can influence responses of "face areas" in certain situations, and even that faces may influence "body areas". Such jointly determined responses may underlie the interaction of face and body information that has been found in behavior. However, as discussed above, human imaging is inherently limited in is ability to probe neural mechanisms. Continuing this line of research with macaque monkeys could reveal if and how face information and body information are integrated at a much finer scale.

In fact, while macaque vision research has generally examined responses to lone faces and lone bodies presented without context, there is one study that provides evidence of interactions between faces and bodies in single cells. Wachsmuth et al. (1994), without any knowledge of the face patches, recorded single neurons along the macaque STS while showing pictures of whole people, lone heads, and lone bodies. They identified 7287 neurons, found just 64 of these that responded to pictures of whole people, and were able to successfully record from 53 of these. While 17 of these cells responded only when the head was present and 5 responded only when the body was present, 22 responded when *either* the head or body were present (but not when inanimate objects were shown). Furthermore, 9 cells responded *only* when *both* head and body were visible. While it is not possible to precisely position these neurons within our current understanding of face and body areas, and although they represent a very small sample, this experiment provides hints of the single neuron

selectivities that may underlie the contextual interactions between faces and bodies.

The macaque face and body patches provide an opportunity to reexamine how bodies impact the representation of faces – and vice versa – within a system that is open to exploration with electrophysiology and causative experiments. Pursuing this opportunity could reveal how viewing bodies influences representations of faces, and more generally how visual context shapes the neural representation of visual objects.

Finding Effects of Visual Context in the Macaque Face-Processing System

To address the questions raised above and identify the cortical areas underpinning the representation of facial context in the macaque, we ran two sets of fMRI experiments: one to explore the processing of facial motion, the other to explore interactions between the representations of face and body form.

In the facial motion experiment, we imaged the brains of macaques while showing them still pictures and movies of geometrical patterns, faces, and non-face objects. This allowed us to carefully compare the intertwined responses to general motion, facial form, and facial motion. In order to examine regions where facial motion might be critical for neural activity, we located an expanded face processing system by comparing responses to moving faces with responses to moving objects. Afterwards, we considered the activity of each face patch when a monkey viewed movies of face motion and temporally scrambled control movies in order to determine which areas – if any – demonstrated selectivity for natural motion. Through this experiment, we found both a previously undiscovered face

patch and a functional division of the face-processing system that suggests that dorsal face patches are specialized for natural face motion.

In the body context experiment, we showed macaques pictures of faces and bodies, alone and together, while imaging activity in the face patches, body patches, and surrounding temporal lobe. This allowed us to study interactions between the neural representations of these two forms. We also used pictures of body-matched objects to examine the form-specificity of observed body context effects, as well as heavily pixelated pictures of faces to probe the underlying meaning of such effects. As we will show, this experiment revealed an unexpectedly supralinear response to faces viewed in the context of bodies in a subset of face patches, a response which was specific to body form and does not appear to result directly from the disambiguating effect of bodies.

These two studies, by identifying areas that demonstrate contextual effects, deepened our understanding of the brain's face-processing machinery and set the stage for future exploration of visual context's neural foundations.

2. Experimental Procedures*

All procedures conformed to local and NIH guidelines, including the NIH Guide for Care and Use of Laboratory Animals. These experiments were performed with the approval of the Institutional Animal Care and Use Committees of The Rockefeller University and Weill Cornell Medical College.

Subjects

We studied four male rhesus monkeys (*Macaca mulatta*), aged 3.5 to 5 years old, weighing 5.5 to 7.5 kg. Here, they are referred to as M1–M4. A cranial implant composed of acrylic cements (C&B-Metabond, Parkell; Palacos LV+G, Zimmer) and anchored with ceramic screws (Thomas Recording; Rogue Research) was implanted in each monkey using standard surgical methods, and standard anesthetic and postoperative treatment protocols were followed. A custom-designed MRI-compatible headpost made of Ultem (SABIC) was secured in each implant.

Data Acquisition

All MRI data were acquired with a 3T Siemens Tim Trio MRI scanner, using an AC-88 gradient insert (Siemens) for functional scans. Immediately prior to each functional scan, a dose of Feraheme (AMAG Pharmaceuticals; M1-M3) or Molday ION (BioPAL; M4) containing 8 to 10 mg/kg of iron was injected into a saphenous vein to increase functional contrast. When Molday ION was administered on consecutive days, we adjusted the dosage to compensate for

---

* Portions this chapter have been published previously (Fisher & Freiwald 2015).

functional half-life. As these agents increase functional contrast by decreasing local MR signal (Mandeville et al. 1998; Vanduffel et al. 2001), we inverted the sign of our measurements during analysis so that positive values reported here reflect increased blood volume. Functional images were obtained with a custom-designed 8-channel surface coil (Lawrence Wald, MGH Martinos Center) and a gradient-echo echo-planar imaging sequence with 54 horizontal slices, a 96 × 96 in-plane matrix, an isotropic resolution of 1 mm^3, TR = 2 s, TE = 16 ms, and 2× GRAPPA acceleration.

We acquired 6 anatomical MRI volumes from each monkey to serve as the basis for individualized models of the cortical surface. With the subject under isoflurane anesthesia and positioned in an MRI-compatible stereotactic frame, T1-weighted images were obtained using a custom single-channel surface coil and a fast gradient echo sequence (magnetization-prepared rapid gradient echo, MPRAGE) with 240 sagittal slices, an in-plane matrix of 256 × 256, and an isotropic resolution of 0.5 mm^3.

Visual Stimuli

During functional scans, stimuli were projected at 60 Hz onto a screen placed 35 cm in front of the monkey's eyes. For the study of facial motion processing (Chapter 3), three block design stimulus sets were presented to each monkey: low-level motion, object category, and object motion stimulus sets (Figure 2). A fourth stimulus set for retinotopic mapping was presented to M1, M2, and M3. For the study of context effects (Chapter 4), an additional body context stimulus set was presented to all 4 monkeys (Figure 3).

Figure 2. Example Stimuli from the Face Motion Study

(**A**) Sample frames from the low-level motion stimulus set and definition of contrast (blue arrow) for measuring low-level motion selectivity.
(**B**) Sample stimuli from the object category stimulus set and definition of the static face selectivity and static body selectivity contrasts.
(**C**) Sample frames from the object motion stimulus set and definition of related contrasts. Note that the source movies used to create the stimuli for the dynamic face selectivity contrast were different than the source movies used for the other blocks of this stimulus set; they featured the same monkeys and cage toys demonstrating different actions. We used data collected in response to the localizer movies *solely* to localize the face patch ROIs, and based all analyses within these ROIs on independent data collected during the presentation of other stimuli.

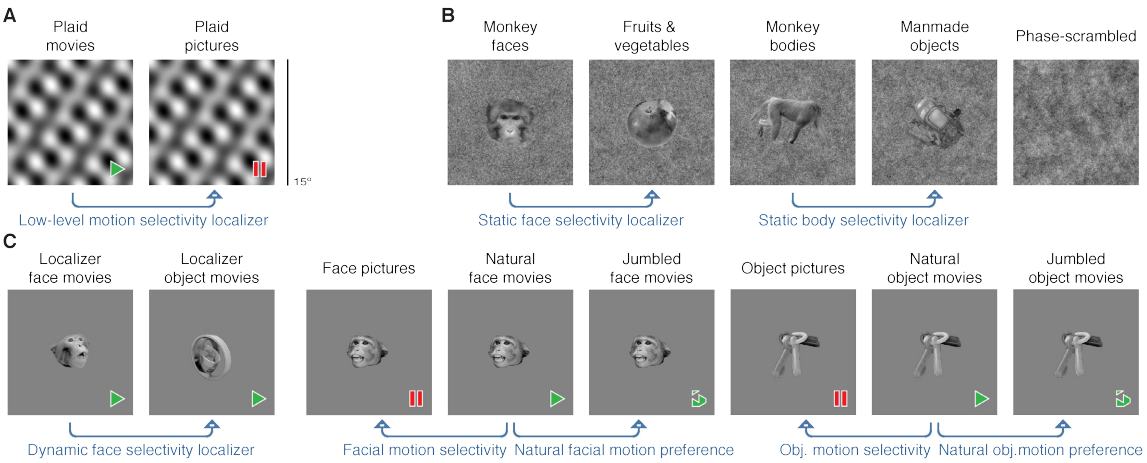
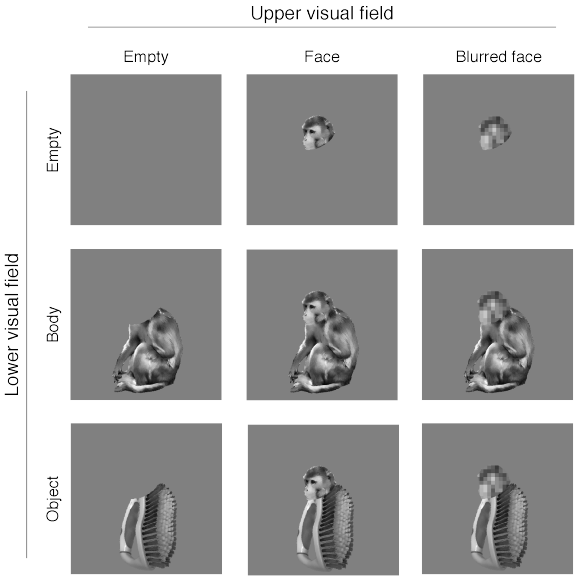

A
Plaid movies
Plaid pictures
15°
Low-level motion selectivity localizer
B
Monkey faces
Fruits & vegetables
Monkey bodies
Manmade objects
Phase-scrambled
Static face selectivity localizer
Static body selectivity localizer
C
Localizer face movies
Localizer object movies
Face pictures
Natural face movies
Jumbled face movies
Object pictures
Natural object movies
Jumbled object movies
Dynamic face selectivity localizer
Facial motion selectivity
Natural facial motion preference
Obj. motion selectivity
Natural obj.motion preference

Figure 3. Example Stimuli from the Body Context Study

Sample stimuli from the body context stimulus set. This is one example monkey/object pair of the 15 that made up this stimulus set. In each case, the object was chosen to match the overall shape and size of the monkey's body. The empty/empty condition was presented in baseline blocks.


Upper visual field
Empty
Face
Blurred face
Lower visual field
Empty
Body
Object

The low-level motion stimulus set (Figure 2A) consisted of two types of stimuli: dynamic and static. Dynamic blocks contained movies of shifting plaids created by superimposing two drifting sinusoidal gratings (Rust et al. 2006). The gratings drifted linearly with a period of 1/3 s and were randomly reoriented after a single period (3 times per second). The static blocks consisted of still frames taken from these movies. The frames of the static stimuli were updated every 1/6 s (every half period) so that 2 frames were shown for each plaid: an initial image and its contrast negative. Thus, while there was a strong percept of translational movement throughout the dynamic blocks, this was removed in the static blocks. These stimuli were presented in the scanner at $15^{o} \times 15^{o}$ and 30 frames per second (fps). A single run of this stimulus set consisted of 36 s of gray, 6 stimulus blocks of 24 s alternating between dynamic and static, 24 s of gray, the 6 stimulus blocks shown in reverse order, and a final 24 s of gray. 2 different orders were used: one where a static block came first and one where a dynamic block came first. 16, 10, 8 and 16 runs of this stimulus set were analyzed, respectively, for M1 to M4.

In the object category stimulus set (Figure 2B), grayscale pictures of rhesus macaque faces, cynomolgus macaque faces, fruits and vegetables, monkey bodies, and manmade objects were presented in separate 24 s blocks. An additional block, composed of phase-scrambled versions of the fruit and vegetable pictures, was also presented. We used this stimulus set to both localize the face patches by comparing the responses to monkey faces with the responses to fruits and vegetables and localize body patches by comparing the responses to monkey bodies with the responses to manmade objects. Both blocks of monkey faces were combined for the identification of face patches. Each block of this

stimulus set consisted of 15 exemplar images shown for 0.4 s, with each exemplar repeated 4 times in a block. Images were matched within and across all categories for total screen area, placed on a pink noise background, and normalized for luminance and frequency amplitude (http://www.mapageweb.umontreal.ca/gosselif/shine/; Willenbockel et al. 2010). In addition, images were chosen so that the aspect ratios of the fruits and vegetables matched those of the macaque faces, and the aspect ratios of the manmade objects matched those of the macaque bodies. At the scanner, the pink noise background was presented at $15^{o} \times 15^{o}$, and the average foreground image was $7.1^{o} \times 5.5^{o}$. Each image was presented at 1 of 5 equally likely screen positions: either at the screen center or offset 1.1^{o} from the screen center to the left, right, top, or bottom. A Latin square design was used to create 6 orders of the stimulus blocks, so that both the relative position of each block within a run and the identity of the immediately preceding block were balanced. Each run of this stimulus set consisted of 36 s of gray, followed by the 6 stimulus blocks without intervening gaps, 24 s of gray, the 6 blocks presented in reverse order, and a final 24 s of gray. In cases where matched sets of all 6 orders were not available for analysis, the palindromic design of the runs allowed balance to be achieved by analyzing matched sets of 3 orders. 36, 30, 27, and 30 runs of this stimulus set were analyzed, respectively, for M1 to M4.

The object motion stimulus set (Figure 2C) consisted of stimuli showing either familiar macaque faces ("faces") or familiar cage toys ("objects") as "pictures", "natural movies", or "jumbled movies". Picture, natural movie, and jumbled movie stimuli were created from a common set of videos: picture blocks showed the videos at 2 fps, natural movie blocks showed them at 15 fps, and

jumbled movie blocks showed the same frames in randomized order at 15 fps. Independent videos were used to create an additional block of natural face movies and an additional block of natural object movies to use for the localization of face patches (in Figure 6A); we refer to these as "localizer movies". Each block of this stimulus set consisted of 6 exemplar movies/image sequences shown for 3 s apiece. Source videos of both monkeys and objects were recorded at 60 fps to minimize motion blur. The cage toys shown in the object blocks were all able to undergo non-rigid movement to better match the range of transformations demonstrated by the monkey faces, and object videos showing vigorous movement were captured to ensure that motion content was no greater in the face videos than in object videos (also, see Figure 8). Stimulus subjects were recorded in front of a blue screen under a standardized set of lighting conditions, monkey bodies were masked with a blue neck plate, and chroma keying was carried out in Final Cut Pro (Apple) so that the subjects could be isolated from the background and videos could be standardized. An exemplar frame from each video was used to standardize the pixel area of the subjects to an equivalent circular diameter, the centers of all subjects' trajectories were aligned, color was removed, and luminance was histogram-normalized across all frames (Willenbockel et al. 2010). At the scanner, the resulting stimuli were shown with the subjects having an equivalent circular diameter 8.5^{o} and moving within a 15^{o} x 15^{o} middle gray background. Again, a Latin squares design was used, resulting in 8 counterbalanced orders. Each experimental run was composed of 24 s of gray, the 8 stimulus blocks (18 s apiece), 24 s of gray, the stimulus blocks in reverse order, and a final 30 s of gray. Palindromic design allowed for balance with matched sets of 4 orders when available runs were not

balanced across all 8 orders. 48, 32, 20 and 40 runs of this stimulus set were analyzed, respectively, for M1 to M4.

The retinotopic mapping stimulus that we showed to M1, M2, and M3 consisted of a bar containing a high contrast moving pattern (the shifting plaids of the low-level motion stimulus) crossing a gray visual field in 8 directions (Dumoulin & Wandell 2008). The circular area traversed by the bar was 30° in diameter at the scanner. This stimulus began with a 30 s period of gray, each crossing of the bar took 32 s, and 20 s periods of gray were placed after every 2 crossings. 19, 31, and 17 runs of this stimulus set were analyzed for M1, M2, and M3, respectively.

The body context stimulus set (Figure 3) consisted of 8 conditions, in which grayscale exemplars of different categories occupied the upper and lower visual fields. The upper visual field could contain either a monkey face ("face") or a heavily pixelated version of the same monkey face ("blurred face"), while the lower visual field could contain either a monkey body ("body") or an inanimate object shaped and sized to match a given body ("object"). The 8 resulting conditions (which were presented in a block design) were therefore "face", "blurred face", "body", "object", "face & body", "face & object", "blurred face & body", and "blurred face & object". Each exemplar face (or its blurred version) was only matched with the body of the same monkey or an object chosen and scaled to match the body. All pictures of monkeys and objects were histogram-normalized (Willenbockel et al. 2010) before being assembled into final stimuli. All stimuli were presented on a middle gray background at 15° x 15° in the scanner. Each block of this stimulus set consisted of 15 exemplar images shown for 0.4 s, with each exemplar repeated 3 times in a block. A Latin squares design

was used to plan experimental runs, resulting in 8 counterbalanced orders. Each run of this stimulus set began with 24 s of grey, the 8 stimulus blocks (18 s apiece), 24 s of grey, the 8 stimulus blocks presented in reverse order, and a final 24 s of grey. As before, palindromic design allowed for balance with matched sets of 4 orders when available runs were not balanced across all 8 orders. 48, 28, 24 and 40 runs of this stimulus set were analyzed, respectively, for M1 to M4.

Data Analysis

We carried out the preprocessing of our fMRI data in a way designed to account for the artifacts most common when imaging of head-fixed, awake monkeys (see discussion in Farivar & Vanduffel 2014). To correct for image changes resulting from head and body motion, slice-wise correction within each run was carried out using AFNI's 3dAllineate (http://afni.nimh.nih.gov/) and included terms for cubic warping in the phase-encoding direction, as well as shifts, rotations, scaling, and skewing within each slice. We then slice-time corrected all runs with FreeSurfer's functional analysis stream (FS-FAST, http://surfer.nmr.mgh.harvard.edu/fswiki/FsFast/; Fischl 2012) and used mutual information-based non-linear alignment (JIP, http://www.nitrc.org /projects/jip/; Mandeville et al. 2011) to compensate for static distortions of the functional volumes and align them to the high-resolution anatomical scans that we used for cortical surface modeling. In order to increase our ability to identify areas showing nonlinear effects of viewing face and bodies together (see Figure 10, purple), we smoothed the data from the body context experiment using a Gaussian kernel with 2 mm full width at half maximum to increase the signal-to-

noise ratio before analysis and mapping. We used unsmoothed data in all other cases, including all other analyses in the body context experiment.

We analyzed the preprocessed data recorded during presentation of the low-level motion, object category, object motion, and body context stimuli with FS-FAST, as described previously (Tsao et al. 2003; Fischl 2012). For retinotopic mapping, we fit a population receptive field model to the recorded data using mrVista (http://vistalab.stanford.edu/software/; Dumoulin & Wandell 2008). As in the FS-FAST analysis, a wide model hemodynamic response function was used to match the expected effect of the contrast agent (Mandeville et al. 1998; Vanduffel et al. 2001).

When mapping functional data, either within brain volumes or onto cortical surface models, we considered two outputs of our fMRI analysis: the percent signal change of a voxel (either between baseline and a stimulus block or between linear combinations of stimulus blocks) and the statistical significance of this percent signal change. Because of inter-subject differences, including MRI coil placement and contrast agent clearance, a given percent signal change does not reliably reflect the same degree of hemodynamic activity in every subject. To compensate for this, signal change was normalized by separately scaling each subject, dividing signal change values of every voxel by the maximal signal change seen in the STS for a contrast of interest. Signals resulting from the low-level motion stimuli were normalized by finding the voxel in the STS with the highest percent signal difference between plaid movies and plaid pictures and dividing all percent signal change values from the same animal by that number. Object category data were normalized by the maximum value of the static face-selectivity localizer within the STS and object motion data were normalized by

the maximum STS value of the dynamic face-selectivity localizer. Because these normalizations were constant multiplicative scaling within each subject, they had no effect on the statistical significance of the maps. Significance values for the maps were adjusted for multiple comparisons by considering all voxels (cortical, subcortical, and cerebellar) within a whole brain mask and using the Benjamini-Hochberg procedure to calculate *q* values representing false discovery rate (Benjamini & Hochberg 1995).

A model of each monkey's cortical surface was created from the anesthetized anatomical scans using FreeSurfer (http://surfer.nmr.mgh.harvard.edu/; Tsao et al. 2003; Fischl 2012). Data (signal changes and *q* values) were projected from voxels in the midway point of the cortical sheet onto cortical surface models using trilinear interpolation. Images of cortical surface models with functional overlays were rendered using custom code in MATLAB (The MathWorks), while images of volume slices with functional overlays and the volumetric reconstruction of ROIs (Figure 6C) were created in Slicer (http://www.slicer.org/).

We analyzed the spatial similarity of the anatomical distributions of static face selectivity and low-level motion selectivity in the fundus of the STS (in Figure 4C-D) by using the overlapping coefficient (hereafter referred to as "distribution overlap"). We restricted our analysis to parts of the STS with a concave curvature and anterior to the MT/MSTv/FST complex, as areas in this complex or along the lips of the STS generally show clear separation of face and low-level motion signals (Polosecki et al. 2013). To identify voxels in the STS fundus that would be likely driven by our stimuli, we subselected from these voxels, choosing the 500 in each subject with the greatest significant signal

change in response to either the localizer face movies or the localizer object movies (Figure 2C); note that responses to these movies were not used to calculate either of the functional maps being compared. We then looked, per voxel, at the selectivity for static faces and low-level motion (normalized per subject) and measured their similarity across the 2000 voxels. Distribution overlap was calculated by normalizing the sum of each contrast over all 2000 voxels to 1, and then summing the lesser of these two normalized distributions on each voxel over all voxels. To determine the significance of this overlap, we used a Monte Carlo permutation test. We created an empirical null distribution (Figure 4D) by randomizing the pairing of the two compared contrasts across the considered voxels within each monkey 1,000,000 times and recording the resulting distribution overlap each time. We calculated the 99% confidence interval of the resulting *p* value according to the binomial distribution (see Ernst 2004). We report the upper (less statistically significant) bound of this confidence interval.

For group analysis, regions of interest (ROIs) were defined in each hemisphere of each animal and combined into bilateral ROIs. Within a given hemisphere, we positioned each ROI by finding a voxel showing a local maximum signal change in an area, selecting a 3 x 3 x 3 voxel region centered on this point, and excluding all voxels that did not meet a $q < 0.05$ significance criterion. If multiple maxima were found which could represent a given area of interest, we chose the one that maximized the symmetry between the hemispheres of a subject. To display the location of these ROIs on cortical surface models, their volumetric locations were projected from voxels at the midway plane of the cortical sheet. In cases where this resulted in a single ROI being

projected to multiple disconnected locations on the cortical surface, only the largest projected location was kept. In some instances, when the included voxels did not include any that fell at the midpoint of the cortical sheet, a volumetric ROI would not appear on the cortical surface model. In these cases, an asterisk was manually placed on the map to denote the approximate location of the unseen ROI.

ROIs were named based on both anatomical and functional characteristics. Face patches were identified using the dynamic face-selectivity localizer and named based on anatomical location and relative position (Tsao, Moeller, & Freiwald 2008). The toy patch was also identified using the dynamic face-selectivity localizer, and was defined by responding more to object movies than face movies. Because there were a number of such areas in each hemisphere (and no existing classification scheme), we chose the most anterior local maximum that was posterior to AL in each hemisphere, as we found that this region reliably neighbored both face patches and areas responding to low-level motion (generally LST). Body patches were identified using the static body-selectivity localizer and named based on anatomical location and relationship to identified face patches (Popivanov et al. 2012); because the currently recognized midSTS and antSTS body patches often include multiple discrete areas of body selectivity, we chose the body-selective activation best-described as between ML and MF to be midSTS, and the one best-described as medial to AL along the ventral bank of the STS to be antSTS. LST was identified as a motion-sensitive area (recognized with the pattern motion localizer) in the lower bank of the STS and anterior to area FST (Nelissen, Vanduffel, & Orban 2006). The MT/MSTv/FST complex (analyzed in Figure 8) was identified as a motion-

sensitive area (recognized with the pattern motion localizer) with appropriate retinotopic organization (M1–M3) and anatomical location (Kolster et al. 2009).

Because the MT/MSTv/FST complex is a functionally heterogeneous region with known retinotopic organization, it was important to define an ROI that spanned this functional diversity and was not overly influenced by a single area or retinotopic preference. The ROI identification scheme used in other regions (a 3 × 3 voxel ROI centered on a local signal maximum) created ROIs that were too spatially limited to meet these criteria. Therefore, we took another approach to define this ROI. We used the low-level motion selectivity localizer (Figure 2A), set a significance threshold with $q << 0.01$ and q small enough that there was no contiguous path of voxels meeting the threshold between the retinotopic center of MT/MSTv/FST and other motion-selective regions. All contiguously connected voxels meeting this threshold were considered part of the MT/MSTv/FST ROI (as in Polosecki et al. 2013).

For group analysis, we independently normalized the time courses from each ROI of each monkey and subjected them to a fixed effects analysis, pooling runs across all monkeys. We did this in two steps. In the first step, the time courses of all voxels within a given monkey and ROI were averaged and the resulting ROI time courses were analyzed with FS-FAST, just as if each ROI was a standard functional voxel. This allowed us to calculate responses within each ROI that could be used for normalization in the second step. In the second step, we normalized the time courses for each ROI in each monkey independently before performing a fixed effects analysis across all monkeys. Note that this is different than the pan-brain per-subject normalization that we performed for the mapping analyses; we chose this alternate approach because differences in signal

distribution (see Janssens et al. 2012) and hemodynamic effects (Handwerker et al. 2012) can occur across the brain – not only between subjects – and it is important to minimize these differences as much as possible prior to a fixed effects analysis. For the object category stimuli, time courses in each face patch were divided by the response to the rhesus faces in that patch. For the object motion stimuli, time courses in each face patch were divided by the response to the localizer face movies in that patch, while time courses in LST and the object patch were normalized by the response to the localizer object movies in that ROI. For the body context experiment, time courses in each face and body patch were divided by the response to whole monkeys (face & body) in that patch. Because each ROI was independently normalized, comparisons between ROIs should be drawn with care, keeping in mind that each signal change is dependent on the response to the condition used for normalization. These normalized ROI time courses were then analyzed across all 4 monkeys using a fixed effects analysis in FS-FAST. We used a fixed effects analysis (as is common in monkey fMRI experiments, e.g., Jastorff et al. 2012; Polosecki et al. 2013) because our sample size was too small for effective power in a random effects analysis (Leibovici & S. Smith 2000) despite our comparatively large number of subjects (see Furl et al. 2012; Polosecki et al. 2013; Janssens et al. 2014). Statistical conclusions drawn from fixed effects analyses generalize only to the sample under study (see discussion in Friston, Holmes, & Worsley 1999). We adjusted the *p* values of ROI analyses for multiple testing using the Holm-Bonferroni method to control for familywise error rate (Holm 1979).

3. Faces in Temporal Context: Responses to Face Motion in the Macaque Temporal Lobe*

We didn't need dialogue. We had faces!

—*Norma Desmond on silent movies,* Sunset Boulevard

Face Motion Activates a Diverse Set of Functionally Specific Areas

Faces carry important emotional, communicational, and identifying information in their movements as well as their form. Face motion activates a wide expanse of cortex in and around the macaque STS (Furl et al. 2012; Polosecki et al. 2013), and this activation may, in part, represent the specific processing of this social signal. However, the degree to which this activation merely reflects underlying sensitivity to general motion or face form remains unclear.

To determine the functional basis of this activity, we identified classical motion areas and face patches for subsequent comparison to the regions that are recruited by face motion. We presented visual stimuli to four rhesus macaque monkeys (M1–M4), and measured the resulting brain activity with fMRI; we used this same technique throughout this study (both Chapter 3 and Chapter 4). Motion areas were localized by contrasting the responses evoked by drifting plaid patterns to those evoked by still plaids (Figure 2A). Face patches were localized by comparing the responses to pictures of faces with responses to fruits and vegetables matched in low-level properties (Figure 2B).

---

* Portions this chapter have been published previously (Fisher & Freiwald 2015).

The selectivity maps for motion and faces (Figure 4A–B) revealed distinct functional specializations within the STS: motion areas MT, MSTv, FST and LST (Nelissen, Vanduffel, & Orban 2006; Kolster et al. 2009), and face patches PL, ML, MF, AL and AF (Tsao, Moeller, & Freiwald 2008). Face patch AM was sometimes difficult to locate, due to chronic contrast agent administration and the subsequent functional signal loss in the ventral brain (Lafer-Sousa & Conway 2013; Gagin et al. 2014). In light of this, we chose to focus our studies on the face patches within or immediately adjoining the STS.

Motion and face activations, despite their proximity in the STS fundus near ML, MF and LST, remained spatially separate: voxels with the highest face-selectivity (Figure 4B, red channel) showed little low-level motion selectivity (Figure 4B, blue; also see Figure 4C–D). Only on the edge of face patches, at the border of motion-selective areas, did concurrent selectivity for low-level motion appear. We found just one exception to this rule of spatial segregation of function: in two subjects the ventral, foveal aspect of the MT/MSTv/FST motion complex exhibited face selectivity (similar to area pPL of Janssens et al. 2014). However, the established face patches are not, as previously suggested (Polosecki et al. 2013), part of a generally motion-selective region within the STS.

Face motion activated some face patches, all identified motion areas, and further outlying areas (Figure 5). Contrasting responses to moving faces with responses to still faces (Figure 2C; Furl et al. 2012; Polosecki et al. 2013) revealed a cortical expanse in and around the STS activated by face motion that included not only face patches, but also all identified motion areas and additional territory in between (Figure 5, left column). Motion of non-face objects, (a contrast comparing object movies to static controls) activated motion areas and anterior

Figure 4. Spatially Dissociated Selectivities for Static Faces and Low-Level Motion in the STS Fundus

(**A**) Left: face form selectivity (red; faces − fruits & vegetables) and low-level motion selectivity (blue; plaid movies − plaid pictures) on M1's right cortical hemisphere. The gray box highlights the region depicted below in B. Right: the hue at each point reflects the relative strength (normalized signal change) of these two contrasts, and opacity reflects the strength of the strongest contrast.
(**B**) Flattened maps, as in panel A, of the STS of each hemisphere in 4 subjects. Face patches (white text) and motion areas (black text) are labeled. Dotted lines represent retinotopic meridians used to identify motion areas and solid lines represent retinotopic meridians at the anterior edge of area V4. Note how there is widespread separation of face selectivity and low-level motion selectivity. We found just one exception to this rule of spatial segregation of function: in M2 and M4 the ventral, foveal aspect of the MT/MSTv/FST motion complex exhibited face selectivity (similar to area pPL of Janssens et al. 2014).
(**C**) Joint face and motion selectivity of voxels in the STS fundus. Static face selectivity and low-level motion selectivity (signal change, normalized per subject) were determined for STS fundus voxels (excluding the MT/MSTv/FST complex; see Chapter 2) most responsive to face or object movies. Color denotes significant ($q < 0.01$) stimulus selectivities of each voxel: red for static face selectivity (faces − fruits and vegetables), blue for low-level motion selectivity (plaid movies − plaid pictures), magenta for both. These selectivity distributions had a voxel-wise distribution overlap value (DOV) of 0.34 (see Chapter 2).
(**D**) Statistical significance of measured DOV. Plot shows the null distribution of DOV, assuming a random association between face-selectivity and motion-selectivity. The red line represents the measured DOV of 0.34 (from panel A). This is significantly ($p < 0.00001$) smaller than expected by chance associations. Thus face- and motion-selectivity are neither co-localized nor independently distributed in this area, but are spatially segregated. Alternative similarity quantifications – cosine distance and Spearman correlation – confirmed the distributions were less similar than expected by chance ($p < 0.0005$). p-values reflect upper (less significant) limits of 99% confidence intervals (see Chapter 2).

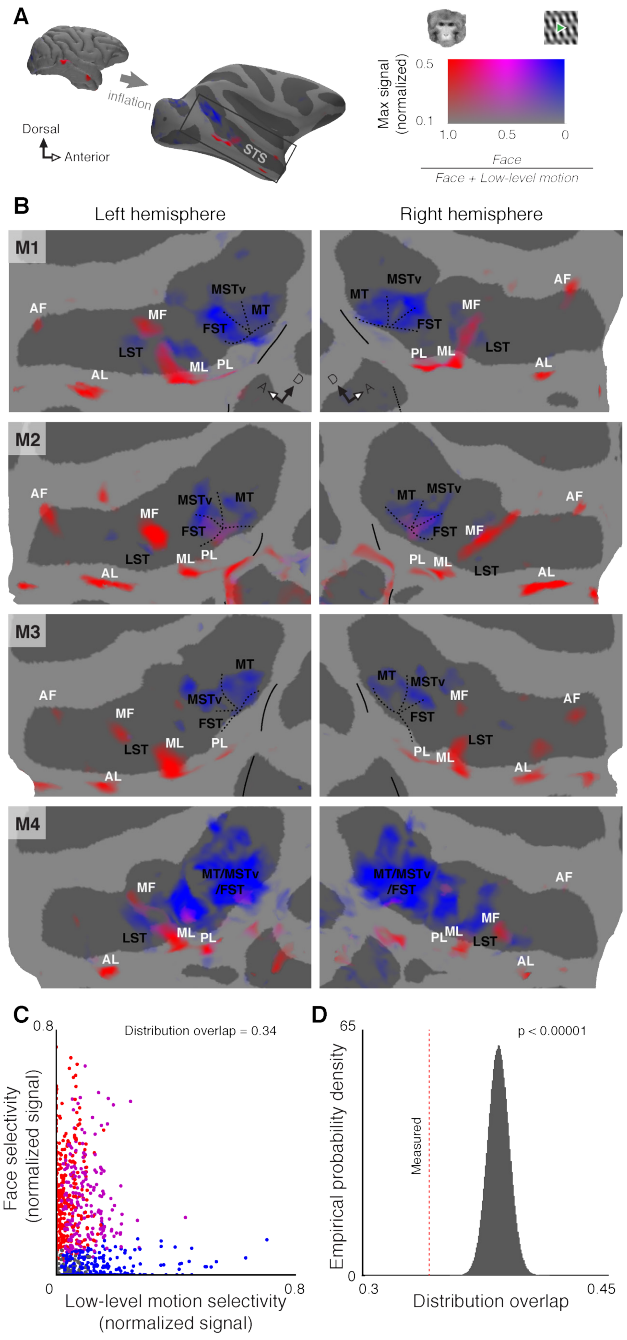
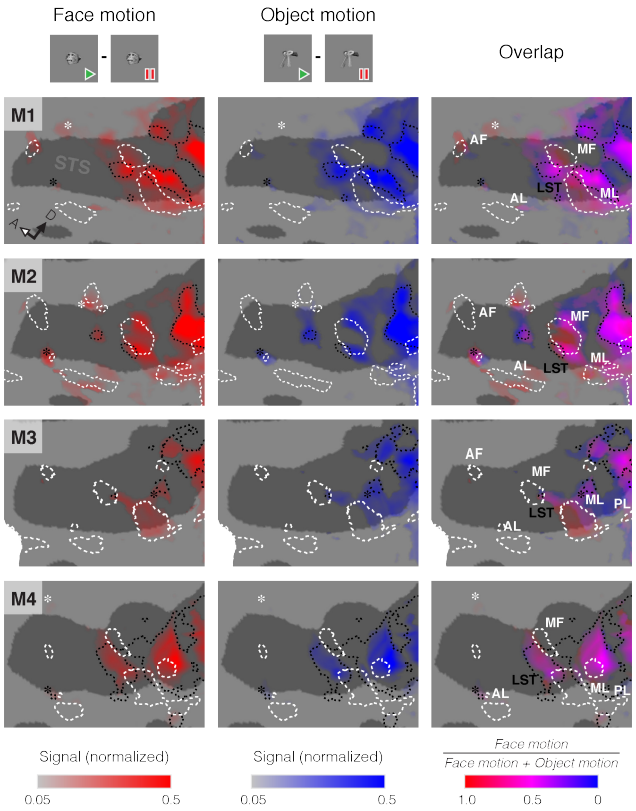

A
Inflation
Dorsal
Anterior
STS
Max signal (normalized)
0.5
0.1
1.0
0.5
0
Face
Face + Low-level motion
B
Left hemisphere
Right hemisphere
M1
M2
M3
M4
AF
MF
MSTv
MT
FST
LST
ML
PL
AL
MT/MSTv /FST
C
Distribution overlap = 0.34
Face selectivity (normalized signal)
0.8
0
Low-level motion selectivity (normalized signal)
D
p < 0.00001
Empirical probability density
65
0
Measured
0.3
0.45
Distribution overlap

Figure 5. Selectivities for Motion Carried by Faces or Non-Face Objects along the Macaque STS

Regions responding to face motion (left column, red; natural face movies − face pictures) or non-face object motion (middle column, blue; natural object movies − object pictures), and the relative strength of these contrasts (right column), in the left hemisphere of each subject. Opacity reflects the contrast strength (normalized signal change). This data is presented on a flattened cortical model of the area surrounding the STS, with dark gray regions representing sulci and light gray regions representing gyri (as in Figure 4B). Dashed white lines outline areas of static face selectivity and dotted black lines outline areas of low-level motion selectivity, both measured in independent fMRI runs (Figure 4). Similarly, white labels indicate face patches and black labels indicate motion areas. Black asterisks highlight areas responding to face and object motion outside of recognized motion processing areas. White asterisks highlight areas more activated by face motion than object motion outside of known face patches. For orientation, the white-filled arrow points anteriorly and the black arrow points dorsally.

Signal change in maps is normalized per-subject and thresholded at a false discovery rate (FDR) of $q < 0.01$.


Face motion
Object motion
Overlap
M1
M2
M3
M4
STS
AF
MF
LST
AL
ML
PL
Signal (normalized)
0.05
0.5
Signal (normalized)
0.05
0.5
Face motion
Face motion + Object motion
1.0
0.5
0

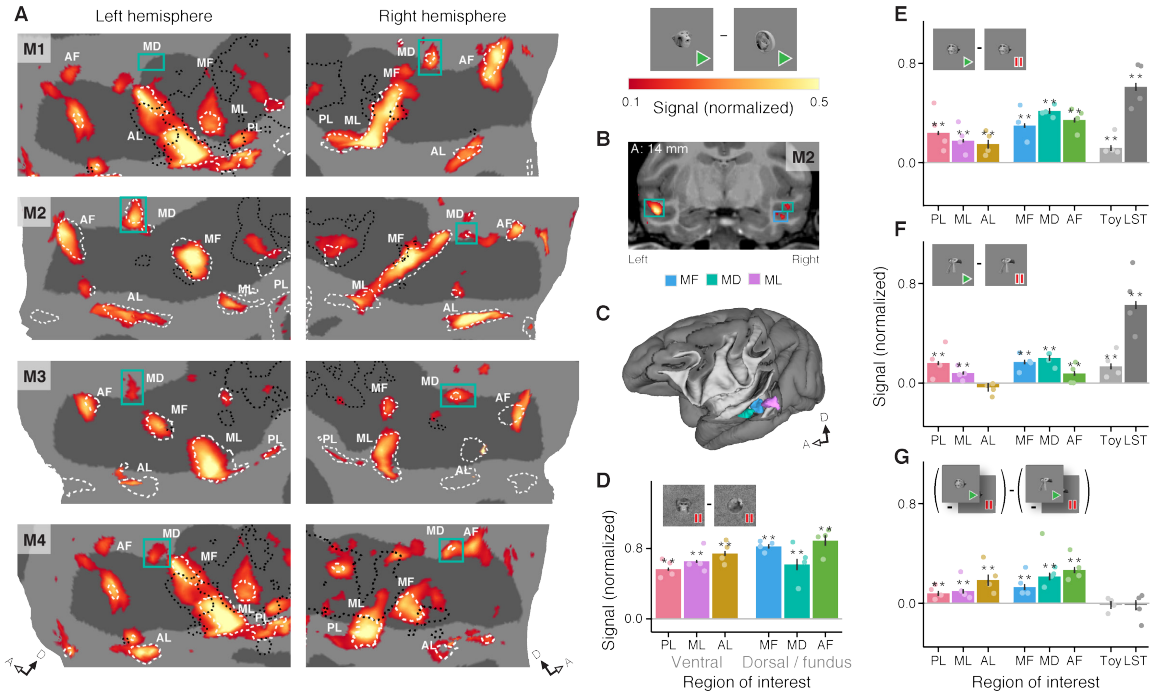
regions (Figure 5, center column) in a pattern largely similar to that of face motion (Figure 5, right column). Some areas responding to the motion of both faces and non-face objects fell outside the general motion areas (Figure 5, black asterisks), suggesting that specializations for forms in motion beyond LST (Nelissen et al. 2011) can be recruited by a range of complex shape motion, including face and non-face object motion. Importantly, there were also regions selectively recruited by face motion but not object motion. These included parts of the known patch face system (Figure 5, white outlines) and additional regions beyond it (Figure 5, white asterisks). These maps show that responses to face motion reach throughout the motion sensitive STS, into at least a subset of face patches, and, intriguingly, beyond the classical face patch system and motion areas.

A Novel Face Patch Responds to Moving Faces

To extend beyond the classical face patch system and map areas that may be particularly attuned to the *motion* of faces (Fox, Iaria, & Barton 2009; Pitcher et al. 2011), we contrasted fMRI responses to movies of faces with responses to movies of articulated toys (Figure 2C). This dynamic localizer (Figure 6A) activated all of the earlier-identified face patches (Figure 4B) and additional parts of the STS's dorsal bank, including many of the areas that had been selectively recruited by face motion (Figure 5, white asterisks). These new dorsal activations included scattered points of face selectivity that varied from individual to individual and, importantly, one area of selectivity at a consistent location in every subject and hemisphere. This area was located anterodorsal to face patches ML and MF

Figure 6. Responses to Complex Motion within an Extended Face Patch System

(**A**) Left: Dynamic face selectivity (localizer face movies − localizer object movies, Figure 2C) on flattened maps of the STS of each hemisphere in 4 subjects. Green boxes highlight the newly described middle dorsal face patch MD, so-called because it is in the dorsal bank of the STS and neighbors middle face patches ML and MF. Dashed white lines outline static face selectivity and dotted black lines outline low-level motion selectivity (as in Figure 5). Right: signal strength color map and schematic of contrast.
(**B**) Coronal slice from M2, showing position of MD and its separation from MF. The right side of the brain is on the right side of the page. The anterior-posterior stereotaxic coordinate is taken relative to the interaural line.
(**C**) Volumetric model of M2's left hemisphere showing the relative locations of ML (purple), MF (blue), and MD (green).
(**D**) Plot of static face selectivity (faces − fruits and vegetables) within the six face patch ROIs that were defined with the dynamic face selectivity localizer (panel A).
(**E** and **F**) Responses to face motion (natural face movies − face pictures) (E) and object motion (natural object movies − object pictures) (F) in the face patches, "toy patch", and LST.
(**G**) Strength of interaction between responses to form and motion: (natural face movies − face pictures) − (natural object movies − object pictures).
* = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm–Bonferroni method for 30 tests (6 ROIs × 1 measure + 8 ROIs × 3 measures). Dots on bar plots represent the values for individual subjects. Error bars represent standard error. Signal change in bar plots is normalized per ROI. Signal change in maps is normalized per-subject and thresholded at an FDR of $q < 0.01$. The raw data analyzed in panels E–G is the same data plotted in Figure 5.


A
Left hemisphere
Right hemisphere
M1
M2
M3
M4
AF
MD
MF
ML
PL
AL
B
A: 14 mm
M2
Left
Right
MF MD ML
C
D
E
F
G
0.1
Signal (normalized)
0.5
0.8
0.0
PL ML AL MF MD AF Toy LST
Signal (normalized)
Ventral
Dorsal / fundus
Region of interest

(Figure 6A), spatially distinct from both (Figures 6B–C, 7). We call this new area the middle dorsal face patch (MD). Thus the pairing of face form and motion reliably recruits six face-specific patches around the STS: PL, ML and AL along its ventral lip, and MF, AF, and the just-recognized MD in its fundus and dorsal bank.

All STS Face Patches Possess a Distinctive Selectivity for Face Motion

The preference for moving faces over moving objects in these six face patches could result from two simpler specializations: selectivity either for face form or for face motion. In fact, all face patches demonstrated similar degrees of selectivity for facial form (Figure 6D) *and* a preference for facial motion (Figure 6E). The facial motion preference was more pronounced in the patches along the fundus and dorsal bank of the STS. Responses to non-face object motion (Figure 6F) were smaller than responses to facial motion throughout. Consistent with this, the interaction between shape category (face vs. object) and motion (moving vs. static) was significant in all STS face patches (Figure 6G). Thus all face patches exhibited a response to motion that was face-specific. Two neighboring control areas, an object-selective STS region that responded more to moving toys than moving faces (referred to as the "toy patch") and motion area LST (Nelissen, Vanduffel, & Orban 2006), were sensitive to both face and object motion to a similar extent (Figure 6E–G). The observed form-specific motion-selectivity of the face patches is, therefore, not due to an imbalance of low-level motion energy across stimuli, a conclusion further supported by balanced activation of general motion areas (Figure 8). Thus selectivity for both the form and motion of faces characterizes all STS face patches, but not the STS at large.

Figure 7 Identification and Location of the Middle Dorsal Face Patch (MD)

(A) Inflated cortical surface model of M2's left hemisphere showing the relative locations of ML (purple), MF (blue), and MD (green).
(B) Slice representations of the middle face patches (ML, MF, and MD) in all subjects (top to bottom). Parasagittal slices (left column) show the left hemispheres. Coronal and horizontal slices (center and right column, respectively) are presented following neurological convention, with the right side of the brain on the right side of the page. Slice coordinates assume standard stereotaxic positioning and are measured from the midpoint of the interaural line. The coronal slice of M2 is the one shown in Figure 6B and is reproduced here for completeness.
Functional color maps show the dynamic face selectivity localizer thresholded at an FDR of $q < 0.01$.

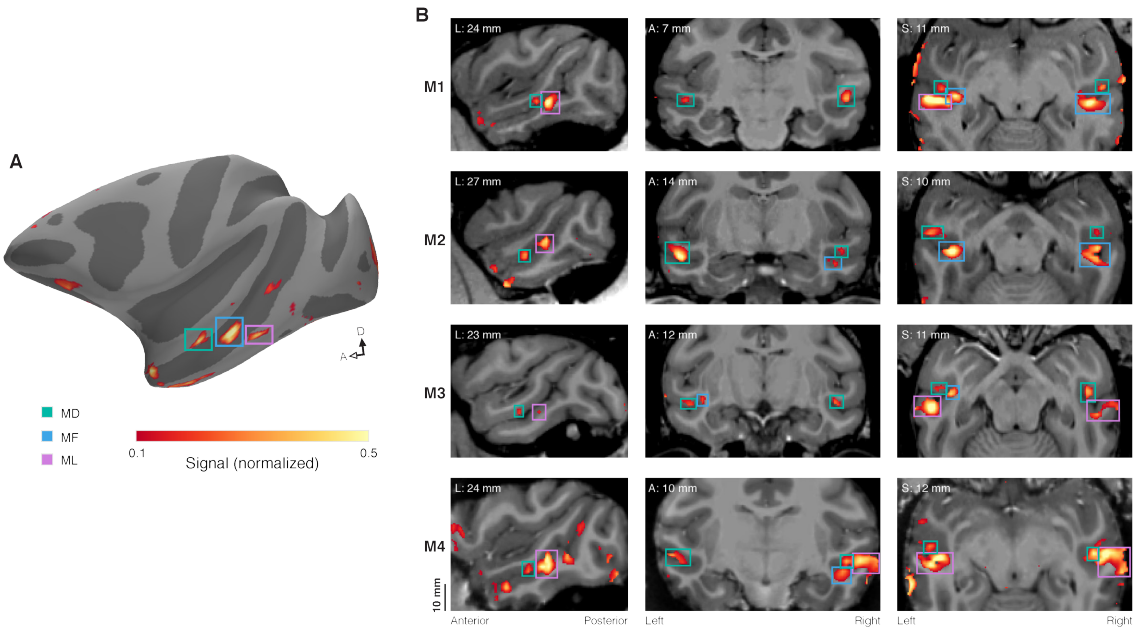
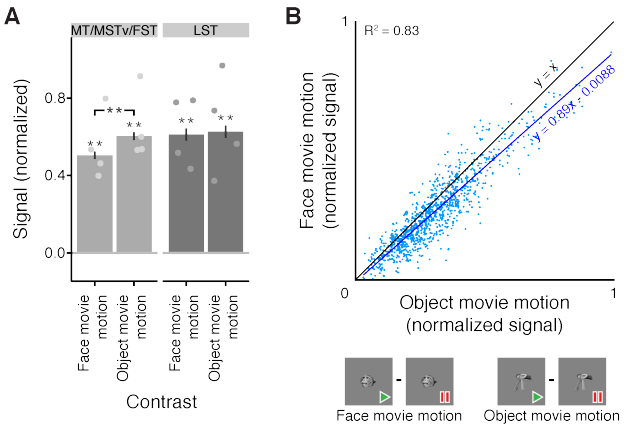

A
D
A
MD
MF
ML
0.1
0.5
Signal (normalized)
B
M1
M2
M3
M4
L: 24 mm
A: 7 mm
S: 11 mm
L: 27 mm
A: 14 mm
S: 10 mm
L: 23 mm
A: 12 mm
S: 11 mm
L: 24 mm
A: 10 mm
S: 12 mm
10 mm
Anterior
Posterior
Left
Right
Left
Right

Figure 8. Analysis of Motion Content in Face and Object Movies

(**A**) Responses of two ROIs, MT/MSTv/FST complex and LST, to two stimulus contrasts, face movie motion (natural face movies − face pictures; "facial motion selectivity" in Figure 2C) and object movie motion (natural object movies − object pictures; "object motion selectivity" in Figure 2C). Both motion-selective regions responded just as much (LST) if not more (MT/MSTv/FST) to the motion in the object movies as to the motion in the face movies. * = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm–Bonferroni method for 6 tests (2 ROIs × (2 measures + comparison)). Dots on bar plots represent the values for individual subjects. Signal change is normalized per ROI. Error bars represent standard error. The data from LST are the same as those presented in Figure 6E–G.
(**B**) Response of the 250 voxels in each subject's temporal lobe most responsive to low-level motion (as measured by the low-level motion selectivity localizer) to face movie and object movie motion. Responses to the two types of motion are highly correlated. Linear regression suggests that the voxels in the temporal lobes most sensitive to low level motion responded ~89% as strongly to motion in the face movies as they responded to motion in the object movies. Signal change is normalized per subject. Both analyses suggest that motion content in face and object movies was well matched, and if not entirely equal, slightly larger in object movies than in face movies.


A
MT/MSTv/FST
LST
Signal (normalized)
0.8
0.4
0.0
**
**
**
**
**
Face movie motion
Object movie motion
Face movie motion
Object movie motion
Contrast
B
R² = 0.83
1
0
1
Face movie motion (normalized signal)
Object movie motion (normalized signal)
y = x
y = 0.89x - 0.0088
Face movie motion
Object movie motion

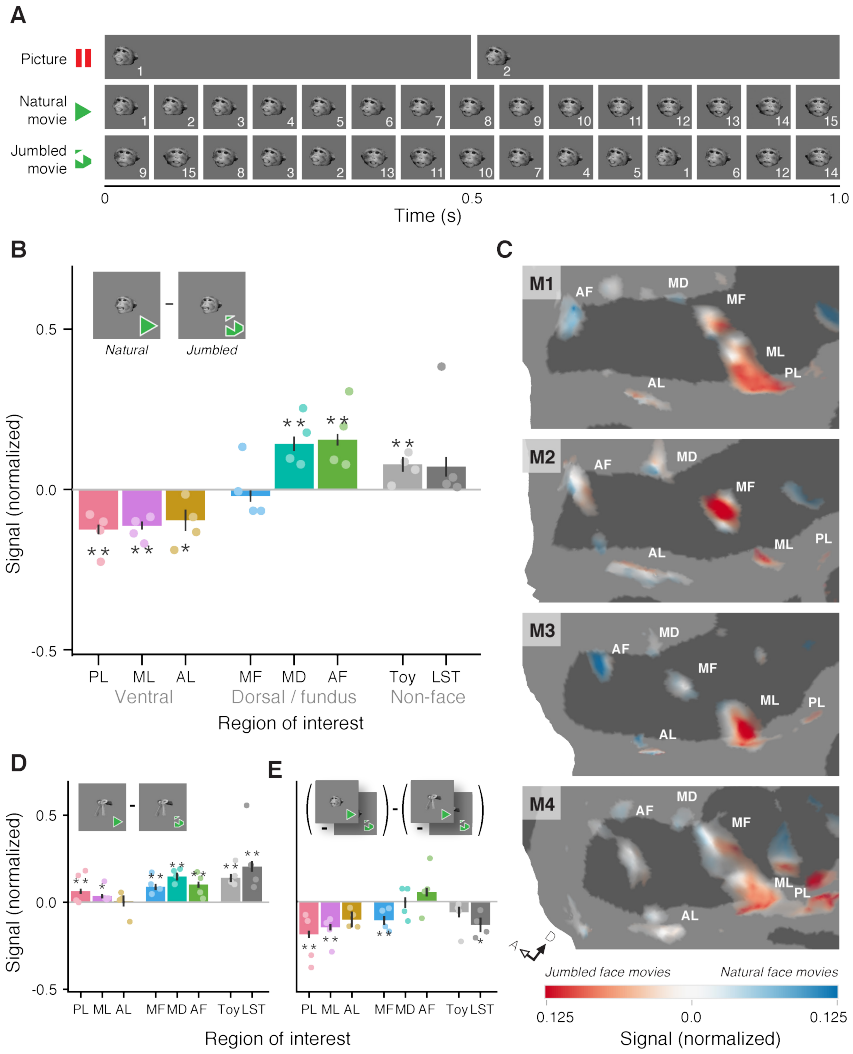
Natural Face Motion Selectivity Divides the STS Face Patch System

We now know that all STS face patches are selective for facial motion (Figure 6G). But does activity within these areas represent *natural* facial motion, or is it simply a response to any update in facial pose, natural or unnatural? We addressed this question by challenging the face-processing system with two stimulus sets that were identical in static content and frame rate, but differed in motion *quality*: the normal, natural movies used earlier in the study, and "jumbled" versions of the same movies, in which frames were presented in a random order (Figures 9A, 2C). We found that dorsal face patches MD and AF showed a significantly greater response to natural movies of faces (Figure 9B). In contrast, ventral face patches PL, ML and AL not only failed to respond more to natural movies, but, surprisingly, their responses were significantly stronger for jumbled movies. MF, positioned between MD and ML, showed no significant preference for either movie type. Thus the face patch system is fundamentally differentiated along a ventrodorsal axis (Figure 9C): the dorsal portion responds preferentially to natural face movements and the ventral portion responds preferentially to facial shapes undergoing rapid, even random, transitions.

The divergent responses of face patches to natural versus jumbled motion did not extend to non-face objects: no patch preferred jumbled object movies to natural ones (Figure 9D). Furthermore, the two control regions responded more to natural object motion (compared to jumbled object motion) than to natural facial motion (compared to jumbled facial motion). As a result, face patches PL, ML, MF and control area LST showed significant interactions between motion quality (natural or jumbled) and form (face or object; Figure 9E).

Figure 9. Preferential Responses to Natural or Disordered Face Motion within the Face Patch System

(**A**) Schematics of stimuli used for analyses of natural motion selectivity. For faces and non-face objects, picture, natural movie, and jumbled movie stimuli were derived from the same 60 fps source videos. Each video was downsampled to 2 fps in the picture condition and 15 fps in the movie conditions. By randomizing the order of each natural movie's frames, a matched jumbled movie was created. Each exemplar stimulus lasted 3 s; a 1 s period is shown here for demonstration.
(**B**) Preference for natural face motion over jumbled face motion across six face patches and two control regions. Dorsal patches MD and AF show a significant preference for natural face motion, while, conversely, ventral patches PL, ML, and AL significantly preferred the rapidly changing jumbled face movies.
(**C**) Preference for either natural face motion (red) or jumbled face motion (blue), as calculated in panel B, across face-selective cortex. Opacity reflects strength of face selectivity (from Figure 6A).
(**D**) Preference for natural object movies over jumbled object movies.
(**E**) Strength of interaction between form (face or object) and frame ordering (natural or jumbled): (natural face movies − jumbled face movies) − (natural object movies − jumbled object movies).
* = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm-Bonferroni method for 24 tests (8 ROIs × 3 measures). Dots on bar plots represent the values for individual subjects. Error bars represent standard error. Signal change in bar plots is normalized per ROI. Signal change in maps is normalized per subject.


A
Picture
Natural movie
Jumbled movie
0
0.5
1.0
Time (s)
B
Natural
Jumbled
0.5
0.0
-0.5
Signal (normalized)
PL ML AL MF MD AF Toy LST
Ventral
Dorsal / fundus
Non-face
Region of interest
C
M1
M2
M3
M4
AF
MD
MF
ML
PL
AL
Jumbled face movies
Natural face movies
0.125
0.0
0.125
Signal (normalized)
D
E
PL ML AL MF MD AF Toy LST
Region of interest

Thus, while natural motion improved localization of an extended face processing system (Figure 6A), and all constituent areas of this system were selective for an interaction of face form and motion (Figure 6G), this shared selectivity arose from two different specializations: the dorsal face patches (MD and AF) genuinely favor natural facial motion, while the ventral face patches (PL, ML and AL) appear to prefer rapidly changing facial pose, regardless of kinematic meaning.

4. Faces in Spatial Context: Specific Interactions Between Heads and Bodies in the Macaque Temporal Lobe

> *All optical illusions which have for result the exhibition of an isolated portion of a live human body, such as a head liberated from the trunk, a bust without a body, or a body without a head, always surprise and interest the spectator.*
>
> — *"Two New Optical Illusions"*, Scientific American, *November 15, 1884*

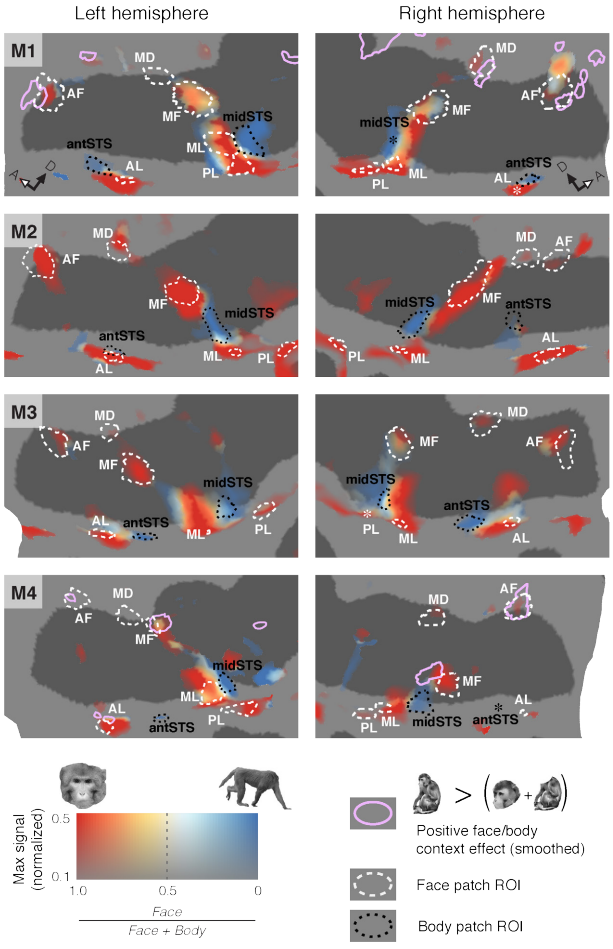
STS Areas Sensitive to the Conjunction of Face and Body

For as long a face patches have been recognized in macaques, body-selective patches have been found nearby (Tsao et al. 2003). Despite these neighboring cortical specializations for two mutually informative biological forms, it is unknown whether the face patches, body patches, or any neighboring STS region integrate information from both faces and their accompanying bodies. Again, we used fMRI to approach this question, and to determine if we could find evidence of such cross-domain contextual effects.

In order to establish the location of the known body patches in the STS and allow for analysis within them, we identified regions that responded more to headless pictures of bodies than to matched pictures of manmade objects (Figure 10, blue), using data from the same fMRI sessions that were used earlier to identify face form-selectivity (Figures 2B). Body-selective areas, though more variable from subject to subject than face patches, were reliably found adjoining face patch ML (often between ML and MF) and medial to face patch AL. These locations agree with the previously described locations of the midSTS body patch

Figure 10. Selectivity for Faces, Bodies, and Their Conjunction in the STS

Flattened maps of the STS in 4 subjects. Face selectivity (faces − fruits & vegetables) and body selectivity (bodies − manmade objects) measured with the object form stimulus set are plotted on each hemisphere; hue at each point reflects the relative strength (normalized signal change) of these two contrasts and opacity reflects the strength of the strongest contrast. Signal change is normalized per-subject and thresholded at an FDR of $q < 0.01$. These two selectivities often neighbor each other, and sometimes overlap (especially evident in M1, M3). Areas that showed a significantly positive response to face/body conjunction (face & body > face + body, $q < 0.05$) were identified using the body context stimulus set and smoothed functional data. These areas are evident in M1 and M4, often overlapping AF, and sometimes overlapping or neighboring AL, MD or MF. The data used to identify these areas are further analyzed in Figures 11–13. Face patch ROIs (white dashes) and body area ROIs (black dots) are labeled according to the criteria described in Chapter 2. In cases where an ROI was not visible on the cortical surface model, an asterisk represents it. The face patch ROIs were defined with the dynamic face selectivity localizer, so that MD could be recognized and so that face patch definitions were constant across the facial motion and body context experiments. Body patch ROIs were defined with the static body selectivity localizer.


Left hemisphere
Right hemisphere
M1
M2
M3
M4
MD
AF
MF
midSTS
antSTS
ML
AL
PL
Max signal (normalized)
0.5
0.1
1.0
0.5
0
Face
Face + Body
Positive face/body context effect (smoothed)
Face patch ROI
Body patch ROI

and the antSTS body patch, respectively (Popivanov et al. 2012). In addition to these reliably identifiable body patches, small areas of body-selectivity were sporadically identifiable anterior to ML (M1, M3 left hemisphere), near MF and MD (M1, M2 left hemisphere, M4 left hemisphere), and adjacent to AF (M1, M3). By independently measuring selectivities for faces and bodies, we could see that these two selectivities sometimes, but not always, overlapped (as in Pinsk et al. 2005; Popivanov et al. 2012).

To search for joint processing of faces and bodies, we scanned each monkey while showing it pictures of whole macaques (head and body), and versions of these pictures where either the head or the body had been isolated (Figure 3). Given the results of previous functional imaging studies (Kanwisher, Stanley, & Harris 1999; Morris, Pelphrey, & McCarthy 2006; MacEvoy & Epstein 2009; Song et al. 2013), it seemed unlikely that we would be able to identify a brain region with an fMRI response to whole monkeys that was stronger than the sum of the responses to isolated heads and isolated bodies; this concern was particularly acute for voxel by voxel whole brain analysis, where low signal-to-noise ratio and multiple comparison correction makes it difficult to identify significant and small signal differences. However, we attempted a whole brain search for evidence of such supralinear responses, using smoothed data to improve the signal-to-noise ratio. We chose to do this so that we might be able to identify regions of selectivity for whole monkeys that could not simply be explained by the superposition of the responses to individual parts, and so that we did not have to limit our search to established functional areas.

Surprisingly, this analysis revealed small areas near some face and body patches that responded more to whole macaques than would have been

predicted by summing the responses to lone faces and lone bodies (Figure 10, purple outlines) in 2 of 4 monkeys (M1 and M4). These areas overlapped with AF in 4 of 8 hemispheres, and neighbored MF in 2 hemispheres (M4), MD in 2 hemispheres (M1), and AL in one hemisphere (M1, left hemisphere). While these regions with a positive context response sometimes fell near the conjunction of face and body patches (as in M4, near MF), at other times they neighbored no regions of identifiable body selectivity (as in M1, near left AF). These maps demonstrate that regions specially attuned to faces and bodies viewed in concert can sometimes be found in the macaque STS near the known face patches.

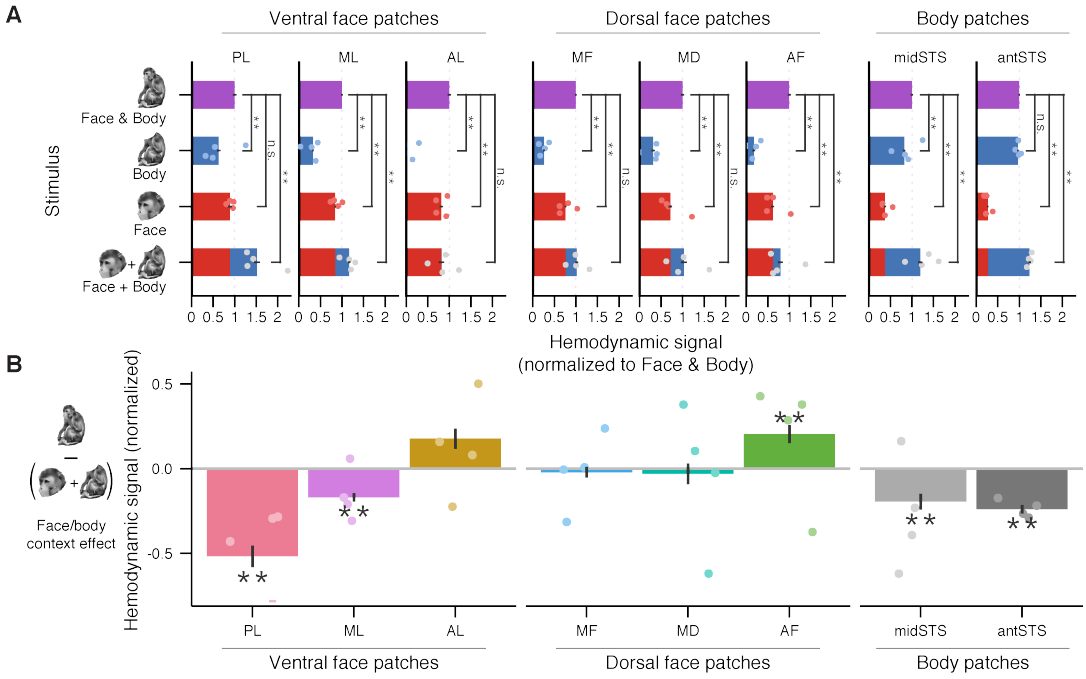
Faces with Bodies Elicit Supralinear Responses in Certain Face Patches

In light of the surprising finding that we could identify supralinear responses to the conjunction of faces and bodies near the face patches, and the psychological evidence that bodies may sometimes affect face-processing mechanisms (e.g. Aviezer et al. 2011; Brandman & Yovel 2012; Rice et al. 2013), we wondered: could positive context responses be localized within the face patches? To address this question, we pooled data across the four monkeys to examine the responses of face patches (Figure 10, dashed white outlines) and body patches (Figure 10, dotted black outlines) to pictures of lone faces, lone bodies, and whole monkeys comprised of both a face and a body (Figure 11).

Although face patches responded more to faces than to bodies and body patches responded more to bodies than to faces, each area generally responded more to pictures of whole monkeys than to its preferred monkey part (Figure 11A). The only exceptions were PL and antSTS, where there was no significant difference between responses to the preferred part and to the whole monkey.

Figure 11. Contextual Effect of Viewing Faces with Bodies

(**A**) Response to lone bodies ("body", blue), lone heads ("face", red), and whole monkeys ("face & body", purple) in 6 face patches and 2 body patches. Two-tone blue and red bars represent the sum of the response to lone bodies and the response to lone heads. All face patches except PL responded significantly more to whole monkeys than to heads alone, and the midSTS body patch responded significantly more to whole monkeys than to bodies alone.
(**B**) The contextual effect of viewing faces in the context of bodies: (face & body) − (face + body). This is the comparison between the purple bar and the red/blue bar in panel A. Face patch AF is activated significantly more by viewing faces and bodies together than by viewing them separately (3 of 4 subjects), while PL, ML, and the body patches are activated significantly less by viewing faces and bodies together. Although AL shows a context effect similar in size to that seen in AF, this effect did not reach significance.
* = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm-Bonferroni method for 88 tests (8 ROIs × 11 measures [across Figures 11–13]). Comparisons labeled n.s. did not reach significance. Dots on bar plots represent the values for individual subjects; dashes represent dots outside of the plotted bounds. Error bars represent standard error. Signal change in bar plots is normalized per ROI to the response to "face & body" stimuli.


A
Ventral face patches
Dorsal face patches
Body patches
PL
ML
AL
MF
MD
AF
midSTS
antSTS
Stimulus
Face & Body
Body
Face
Face + Body
n.s
0 0.5 1 1.5 2
Hemodynamic signal
(normalized to Face & Body)
B
Face/body
context effect
Hemodynamic signal (normalized)
0.5
0.0
-0.5
PL
ML
AL
MF
MD
AF
midSTS
antSTS
Ventral face patches
Dorsal face patches
Body patches

While this arguably shows a benefit for whole monkeys over faces in most of the face patches (and for whole monkeys over bodies in midSTS), this benefit could derive from independent selectivities for faces and bodies rather than joint processing of the whole animal.

When comparing responses to whole monkeys with the sum of responses to lone faces and lone bodies, we discovered qualitatively different results in different patches (Figure 11B). PL, ML, and both body patches responded less to whole monkeys than summing the responses to individual parts would suggest. In fact, summing the responses to lone faces and bodies in PL predicted a response to whole monkeys 51.9% (95% CI: [39.4%, 64.4%]) greater than the one that was measured. AF, on the other hand, responded *more* to the pictures of whole monkeys than a linear accounting predicted (in 3 of 4 monkeys). The response to monkey faces in this patch only accounted for an average of 61.7% (95% CI: [55.2%, 68.3%]) of the response to whole monkeys; with the body response accounting for 17.8% (95% CI: [11.3%, 24.3%]) of the activity, 20.5% (95% CI: [10.0%, 30.1%]) of AF's response to whole monkeys could be considered a positive contrast effect. The contrast effect in AL was similar in magnitude (17.6%; 95% CI: [6.0%, 29.5%]), but did not reach our threshold for statistical significance after correction for multiple comparisons. This shows a functional diversity within the face patch system, with posterior patches showing sublinear responses to faces and bodies viewed together while AF (and possibly AL) showing a supralinear response to faces viewed in the context of bodies.

The Form-Specific Contextual Effects of Viewing Faces with Bodies

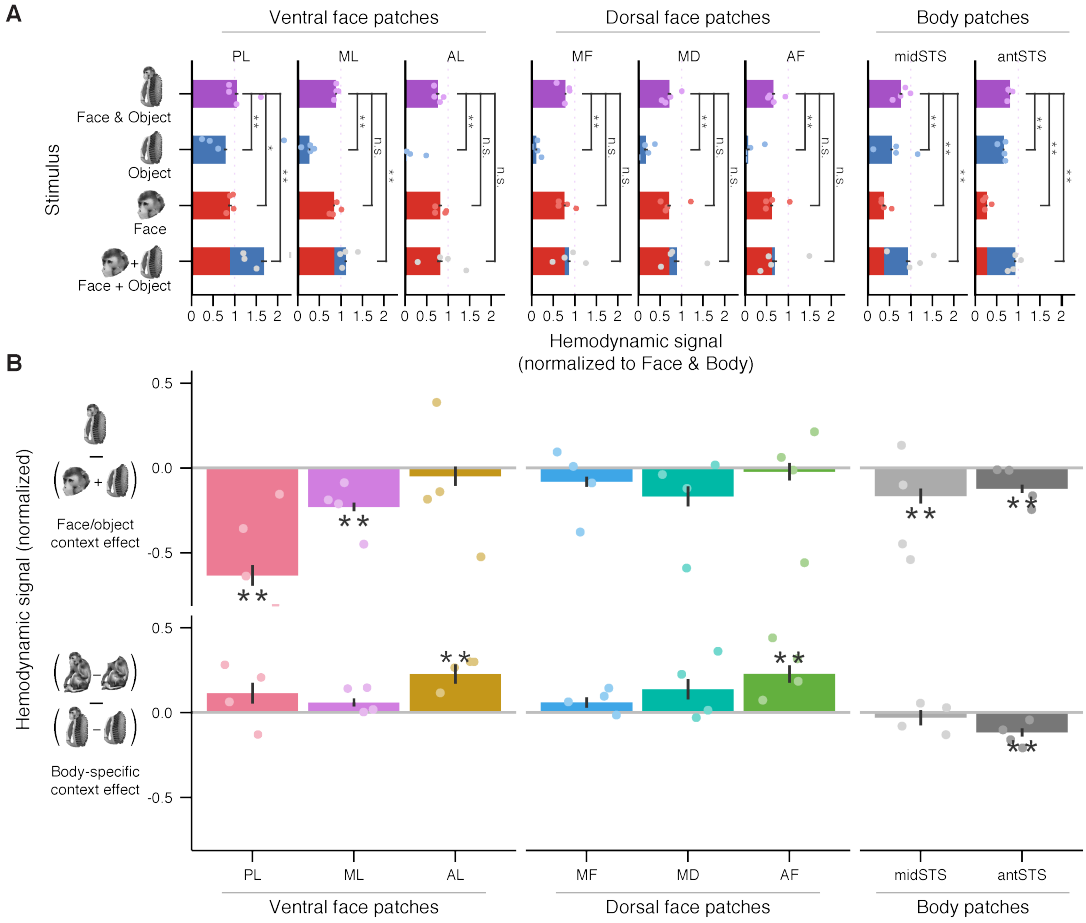
Although AF showed a clear increase in response when a face was viewed on top of a body, it was unclear how important the specific form of the body was for eliciting this benefit. Therefore, we next examined how each functional area responded when a face was shown not with its accompanying body, but instead with an inanimate object that had been chosen to match the overall shape and scale of the body (Figure 3, bottom row).

In contrast to what we saw when faces were presented with bodies, viewing faces with non-body objects did not significantly increase the response of most face patches beyond the response elicited by viewing faces alone (Figure 12A). Again, PL was the lone exception. Relatedly, there were no face areas that showed a significantly supralinear response when faces were viewed with non-body objects (Figure 12B, top). Unsurprisingly, perhaps, this was also true of the body areas.

To determine the body-specific context effect within each face patch, we subtracted the context effect that was elicited by pairing faces with objects from the context effect that was elicited by pairing faces with bodies. This comparison has the benefit of accounting for any low-level retinotopic effects that occur when a body is shown along with a face. We found that both AL and AF showed positive body-specific positive context effects (4 of 4 monkeys; Figure 12B, bottom). The size of this effect was 22.7% (95% CI: [11.4, 33.9]) of the overall response to a whole monkey in AL, and 22.8% (95% CI: [12.7, 32.9]) of this response in AF. Body patch antSTS was the only area to show an increased negative context effect when faces were paired with bodies rather than objects.

Figure 12. Effect of Viewing Faces in the Context of Non-Body Objects

(**A**) Response to non-body objects ("object", blue), macaque heads ("face", red), and their conjunction ("face & object", purple) in 6 face patches and 2 body patches. Two-tone blue and red bars represent the sum of response to objects and the response to heads. No face patches, save PL, responded significantly more to heads presented on top of objects than to heads alone.
(**B**) Top: the contextual effect of viewing faces paired with non-body objects, i.e., (face & object) − (face + object). This is the comparison between the purple bar and the red/blue bar in panel A. No face or body patch was activated significantly more by viewing faces and objects together than by viewing them separately. Bottom: body-specific contextual effect of viewing faces with bodies, i.e., ((face & body) − body) − ((face & object) − object). This is the difference between Figure 11B and the top plot in this panel. Both AL and AF show significantly more benefit when viewing a face with a body than when viewing a face with a matched non-body object.
* = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm-Bonferroni method for 88 tests (8 ROIs × 11 measures [across Figures 11–13]). Comparisons labeled n.s. did not reach significance. Dots on bar plots represent the values for individual subjects; dashes represent dots outside of the plotted bounds. Error bars represent standard error. Signal change in bar plots is normalized per ROI to the response to "face & body" stimuli.


A
Ventral face patches
Dorsal face patches
Body patches
PL
ML
AL
MF
MD
AF
midSTS
antSTS
Stimulus
Face & Object
Object
Face
Face + Object
Hemodynamic signal
(normalized to Face & Body)
B
Hemodynamic signal (normalized)
Face/object
context effect
Body-specific
context effect
Ventral face patches
Dorsal face patches
Body patches

The negative context responses that PL and ML showed with bodies were not significantly different from the negative context responses that they showed with non-body objects, suggesting that these were form-agnostic competition effects. This comparison shows that while some spatial contextual effects in the face patches are not form-specific, body-specific context effects in anterior face patches AL and AF represent a sizable portion of the activity elicited by viewing a whole monkey.

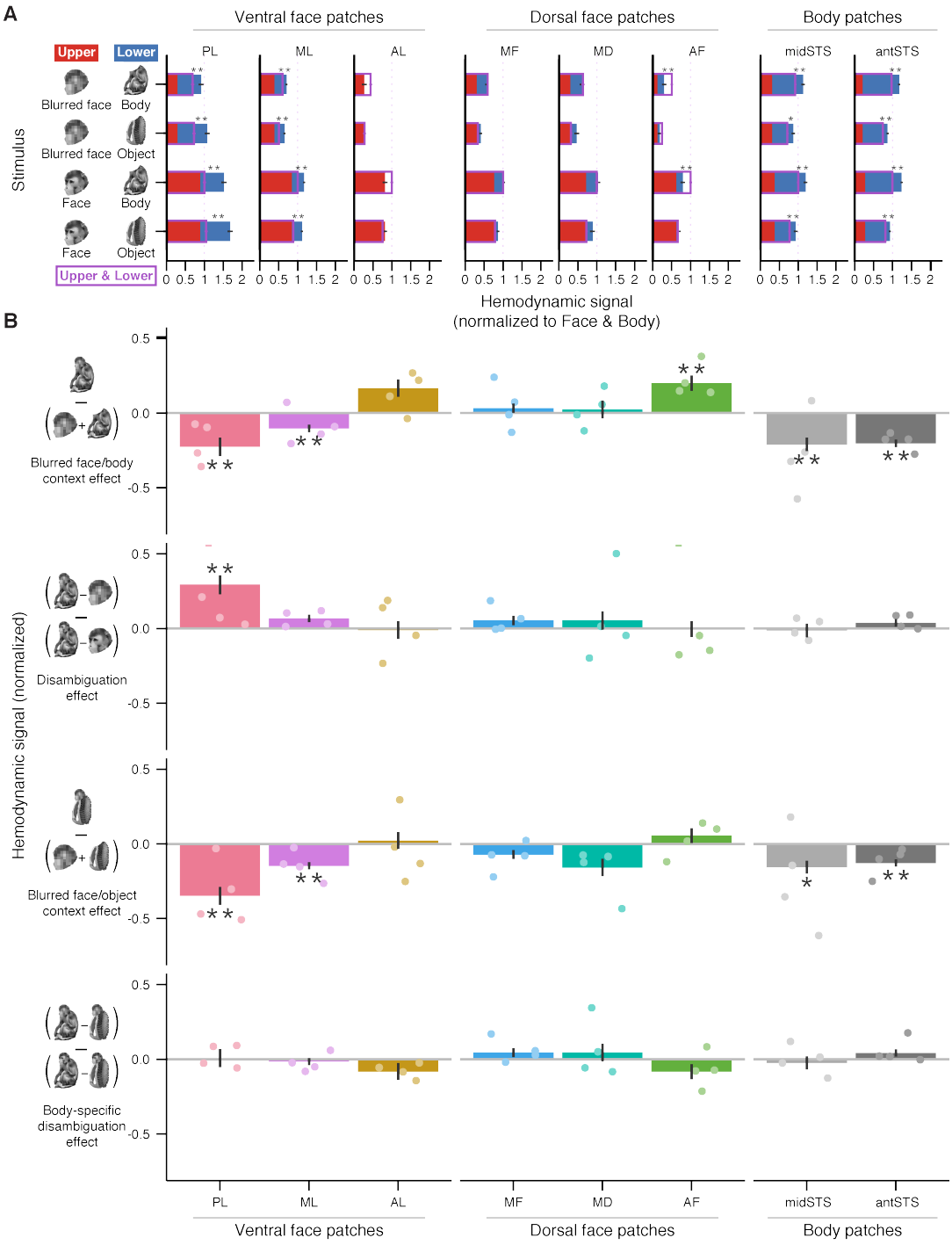
Faces and Ambiguous Faces Elicit Similar Body Context Effects

Given the hypothesis that some of the contextual effects of pairing heads with bodies results from bodies providing evidence that ambiguous faces are, in fact, faces (Cox, Meyers, & Sinha 2004; Brandman & Yovel 2012), we decided to investigate whether the effects we were seeing were the result of such disambiguation. If this were the case, we believed that we should be able to increase the strength of the context effects we observed, or elicit context effects in additional face patches, by increasing the ambiguity of the faces in our stimuli. We examined this possibility by showing the monkeys versions of the stimuli in which internal details of the faces were blurred by heavy pixelation (Figure 3, right column).

As expected, the face patches responded less when faces were blurred than when faces were unobscured (Figure 13A), though interestingly, they still showed a contextual effect with blurred faces (Figure 13B, top). AF showed a significant positive context effect with blurred faces, as it had with unambiguous faces, responding more to blurred faces on a body than a linear accounting would predict. PL, ML, and the body patches showed a negative context effect

Figure 13. Response to Viewing Ambiguous Faces in the Context of Bodies and Non-Body Objects

(**A**) Effects of viewing 4 different pairs of upper/lower visual field stimuli, either singly or in concert, in 6 face patches and 2 body patches. Responses to lone upper field stimuli (faces or blurred faces) are represented by the red portion of each two-tone bar and responses to lone lower field stimuli (bodies or objects) are represented by the blue portion. Reponses to viewing the pairs together are represented by the purple outlines. Asterisks mark cases where there is a significant difference between the response to viewing upper and lower field stimuli together and the sum of the responses to viewing them separately. The bottom two rows show data from Figures 10 and 11 for the sake of comparison. (**B**) Top: the benefit of viewing ambiguous faces in the context of bodies, i.e. (blurred face & body) − (blurred face + body). This is the comparison tested in the top row of panel A. As was true with unambiguous faces, AF is activated significantly more by viewing blurred faces and bodies together than by viewing them separately. 2nd row: effect of viewing faces on bodies that can be attributed to facial disambiguation, i.e., ((blurred face & body) − blurred face) − ((face & body) − face). This is the difference between the top plot in this panel and Figure 10B. Only PL shows a significantly larger contextual benefit with blurred faces, and it results from less of a *negative* context effect, rather than more of a positive one. 3rd row: the contextual effect of viewing blurred faces paired with non-body objects, i.e., (blurred face & object) − (blurred face + object). No face or body patch was activated significantly more by viewing blurred faces and objects together than by viewing them separately. Bottom: *body-specific* effect of viewing faces on bodies that can be attributed to facial disambiguation, i.e., ((blurred face & body) − (blurred face & object)) − ((face & body) − (face & object)). This effect did not reach significance in any of the examined areas.
* = $p < 0.05$ and ** = $p < 0.01$, corrected using Holm-Bonferroni method for 88 tests (8 ROIs × 11 measures [across Figures 11–13]). Comparisons labeled n.s. did not reach significance. Dots represent single subject values; dashes represent dots outside of plotted bounds. Error bars represent standard error. Signal change in bar plots is normalized per ROI to the response to "face & body" stimuli.


A
Ventral face patches
Dorsal face patches
Body patches
Upper
Lower
PL
ML
AL
MF
MD
AF
midSTS
antSTS
Blurred face
Body
Blurred face
Object
Face
Body
Face
Object
Upper & Lower
Stimulus
0 0.5 1 1.5 2
Hemodynamic signal
(normalized to Face & Body)
B
Blurred face/body
context effect
Disambiguation
effect
Blurred face/object
context effect
Body-specific
disambiguation
effect
Hemodynamic signal (normalized)
0.5
0.0
-0.5
PL
ML
AL
MF
MD
AF
midSTS
antSTS
Ventral face patches
Dorsal face patches
Body patches

with blurred faces. AL, once more, showed a positive context effect that did not reach significance. By subtracting the contextual effect for clear faces (Figure 11B) from the effect for ambiguous faces, it became evident that the contextual effects of bodies on blurred faces were not significantly different from their effects on unambiguous faces in any area other than PL (Figure 13B, 2nd row). When looking at the body-specific context effects, taking into account responses to blurred faces seen atop non-body objects (Figure 13B, 3rd row), even PL showed no difference in context effect between the blurred and unambiguous faces (Figure 13B, bottom). Therefore, bodies shown with either unambiguous *or* ambiguous faces can elicit a supralinear context effect within certain STS face patches, suggesting joint processing of faces and bodies. However, the size of the two context effects is almost identical, arguing against disambiguation as the primary impetus of these effects.

5. Discussion*

Constructing a coherent visual world is not easy. Piecing together faces and extracting the crucial information that they convey – who she is, how he feels, where they are looking – is a challenging undertaking, and it is logical to use every available resource to confront it. Our work shows that the macaque brain, following this logic, may take a catholic approach to ease the burden of understanding faces: the activity of the face patches is dictated not only by facial form, but also by the temporal and spatial context in which faces are seen.

Context-evoked responses are not evident throughout the face processing system, but only in a subset of patches. One set of face patches responds preferentially to naturally moving faces, and another partially overlapping set is particularly activated by faces seen in the context of bodies. Taking the contextual effects within these areas one at a time for the moment, we can consider what they tell us about the organization of the macaque temporal lobe, how they relate to human physiology and psychology, and what our functional imaging results suggest about underlying neural processes.

Localizing Responses to Facial Motion

From just a glance at a face, we gather an abundance of social information (Willis & Todorov 2006). Set in motion, the face comes alive, augmenting this knowledge (B. Knight & A. Johnston 1997; Lander, Christie, & Bruce 1999), but also posing a challenge for the neural systems that must now interpret an evolving subject (Sinha 2011). Our investigation of facial motion processing

* Portions this chapter have been published previously (Fisher & Freiwald 2015).

aimed to identify the neural machinery that navigates these intertwined opportunities and challenges of facial motion, leveraging a model system that is similar to the human face-processing system (Tsao, Moeller, & Freiwald 2008; Yovel & Freiwald 2013), remains highly reproducible across subjects (Tsao, Moeller, & Freiwald 2008), and enables mechanistic exploration of the computations underlying face recognition (Freiwald & Tsao 2010). The specialized areas that we recruited with naturally moving faces may mark a key component of the machinery for dynamic face recognition.

The architecture of face motion processing revealed in the macaque STS includes areas selective for low- and high-level motion (Furl et al. 2012), face form (Polosecki et al. 2013), and natural facial motion. These areas all neighbor each other but remain spatially distinct. This picture of a functionally heterogeneous mosaic represents a fundamental departure from earlier fMRI studies (Furl et al. 2012; Polosecki et al. 2013) which suggested that any motion responsiveness found in dorsal face patches (Polosecki et al. 2013) was a byproduct of these areas overlapping a generally motion responsive region. Our results reflect a different reality: while some STS regions are broadly motion-sensitive – responding similarly to face motion and non-face motion – neighboring areas specifically process *face* motion.

One such area is MD, a newly described face patch in the upper bank of the STS (Figures 6A–C, 7). While MD is occasionally evident when static stimuli are used for mapping (similar to aMF of Janssens et al. 2014), dynamic stimuli allowed us to locate this area in all eight hemispheres that we studied. This pattern is reminiscent of the human pSTS face area, which is critical for processing moving faces (Pitcher, Duchaine, & Walsh 2014), is identified

sporadically with static stimuli but reliably with dynamic ones (Fox, Iaria, & Barton 2009; Polosecki et al. 2013) and shows selectivity for natural face motion (Schultz et al. 2013). Interestingly, the human pSTS face area does not appear to be strongly connected to the ventrally located fusiform and occipital face areas (Gschwind et al. 2012). Similarly in macaque monkeys, when connectivity of face patches was mapped, no strong projections to the location of MD were reported (Moeller, Freiwald, & Tsao 2008). Furthermore, anatomically variable activations by faces are found anterior to both MD (this study and Janssens et al. 2014) and human pSTS (Pitcher et al. 2011). One plausible explanation of this variability is that these anterior regions represent a variety of social signals of diverse complexity (Keysers & Perrett 2004; Hein & R. T. Knight 2008), and are only partially and erratically activated by faces. Thus functional specialization, connectivity, and relative location indicate that MD might be the macaque homolog of the human pSTS face area, and could therefore be critical for establishing general homology between face processing systems of humans and macaques (Tsao, Moeller, & Freiwald 2008; Yovel & Freiwald 2013).

We found a new functional differentiation within the macaque face-processing system in which dorsal patches preferred naturally moving faces, while ventral patches (to our surprise) preferred random transitions in face pose (Figure 9B–C). This reveals a novel dimension of the cortical representation of faces and marks, to our knowledge, the first time that fMRI has revealed an overt functional dissociation – where different areas have significant and opposing selectivities – within the macaque face-processing system.

While it is not currently possible to link changes in fMRI signals to underlying neural activity with absolute certainty (see discussion below), this

preference for natural facial motion suggests that cells in MD and AF could also exhibit selectivity for the kinematics of naturally moving faces in addition to their selectivity for static facial form (McMahon et al. 2014). Some neurons in these patches may fire only in response to a specific sequence of poses, a mechanism that has been proposed for the neural coding of biological motion (Giese & Poggio 2003; Vangeneugden et al. 2011).

On the other hand, the apparent selectivity of ventral face patches PL, ML and AL for randomized face motion is unlikely to represent a genuine selectivity for specific sequences of facial pose. Rather, this preference may reflect purely shape-selective face neurons that adapt quickly (Grill-Spector & Malach 2001), respond less to expected stimuli (Summerfield et al. 2008; Meyer & Olson 2011), or show a combination of these effects (Perrett et al. 2009; Larsson & A. T. Smith 2012). Thus a predictive coding scheme, where deviations from expectation drive neural activity (Friston 2005; Clark 2013), could underlie processing in the ventral patches. While predictive coding models generally assume predictions from later processing levels inform earlier processing levels (e.g. Friston 2005), our discovery of qualitatively distinct representations of facial motion within the face patch system allows an alternative hypothesis to be explored: dynamic face representations in dorsal face patches might generate predictions of momentary features which are communicated to ventral patches through *lateral* connections (Moeller, Freiwald, & Tsao 2008).

The fact that the natural motion benefit in the dynamic dorsal patches was just as great for the object stimuli as for the face stimuli (Figure 9D–E) is difficult to explain at the moment. It could mean that these areas have independent selectivities for face form and for natural, continuous motion; however, the

relative insensitivity of the dorsal STS to the pattern motion localizer speaks against this. Another possibility is that there is a greater difference in natural motion content between our veridical and jumbled object movies than between our veridical and jumbled face movies. The fact that LST shows greater benefit for natural motion with objects than with faces (Figure 9C–E) supports this idea. Unfortunately, there is no way to control for such potential differences without knowing what it is about natural motion that these regions prefer. Experiments that use a wider range of stimuli to dissect the relevant components of motion "naturalness" will likely be necessary to explain this result.

When interpreting this experiment's results, it is also crucial to note that the neural correlates of fMRI signals are a subject of ongoing debate, and links between the signals we describe and specific neural activity are still conjectural. Studies over the past 15 years have variously identified either the integration of neural inputs (Logothetis et al. 2001; Goense & Logothetis 2008) or the spiking of local neurons (D.-S. Kim et al. 2004; Lee et al. 2010; Issa, Papanastassiou, & DiCarlo 2013) as the best correlate of fMRI signals. It is likely that both of these influences and others play a role in generating fMRI signals some of the time (reviewed in S. G. Kim & Ogawa 2012). Furthermore, our contrast-enhanced fMRI technique measured cerebral blood volume (CBV), not the blood-oxygen-level-dependent (BOLD) fMRI signal that has been most frequently studied. While CBV changes are similar to BOLD in most cases, they are decoupled in others (Smirnakis et al. 2007; Goense, Merkle, & Logothetis 2012), so much of what is known about the neural correlates of fMRI signals may not directly apply to our study. Despite this, we have good reason to believe that CBV corresponds well to the firing rate in and around the face patches; Issa et al.

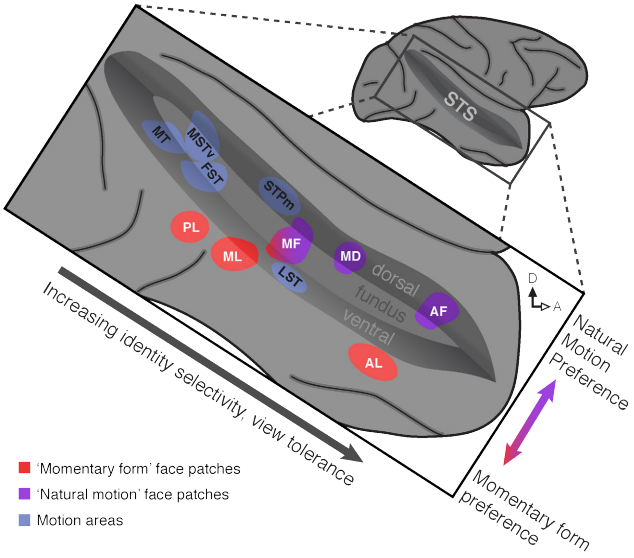
(2013), who also measured CBV, found that a spatially smoothed map of raw spike rates was a very good predictor of the fMRI signals that they measured around the macaque STS. While this suggests that our measured fMRI responses reflect local firing rates, and therefore that natural motion increases the firing rate of dorsal face patch neurons, direct neural recordings are necessary to confirm this.

Furthermore, while our use of jumbled frames as a control revealed a functional divide within the face-patch system, jumbling is a coarse manipulation that introduces discontinuities into continuous motion and interrupts the possible expectation of preserved stimulus identity. This study, therefore, speaks specifically to functional specializations for *continuous* face motion. A recent experiment demonstrated that certain human face areas respond differentially to movies of facial expressions played either forwards in time (a continuous, biologically *plausible* motion) or backwards (a continuous but implausible motion; Reinl & Bartels 2014). A similar comparison in monkeys might deepen our understanding, showing further motion specialization within the face patches or refining the mechanistic understanding of the division we describe.

By integrating our results with the findings of earlier studies, we can develop a picture of how face form and motion processing are arranged in the macaque temporal lobe (Figure 14). Within and around the STS, face patches and general motion areas adjoin each other, but are anatomically distinct. The face patches are differentiated along two axes. As information flows from posterior to anterior, face patches show increased form-specificity and view-tolerance (Freiwald & Tsao 2010), consistent with general trends in the temporal lobe

Figure 14. Model of Face Motion and Face Form Processing along the Macaque Temporal Lobe

Functional specificity of face patches is organized along two main anatomical axes. From posterior to anterior, face patches show increasing identity selectivity and increasing tolerance to viewing condition (Freiwald & Tsao 2010). Along the dorso-ventral axis, face patches show differential selectivity for natural motion, with "dynamic" dorsal patches (purple) responding to natural motion and "static" ventral patches (red) responding more to rapidly varying face stimuli. Face motion activates all of these patches as well as motion processing areas (blue), which are selective for neither momentary face form nor natural face motion.


STS
MT
MSTV
FST
STPm
PL
ML
MF
MD
LST
dorsal
fundus
ventral
AF
AL
D
A
Natural Motion Preference
Momentary form preference
Increasing identity selectivity, view tolerance
'Momentary form' face patches
'Natural motion' face patches
Motion areas

(Rust & DiCarlo 2010). Along the ventral to dorsal axis, there is a functional transition that reflects a likely selectivity for momentary facial form in the ventral patches and for continuous facial motion in the dorsal ones. This picture is compatible with influential "division of labor" face recognition models (e.g. Bruce & A. W. Young 1986), particularly those that posit a separation of static features (such as identity) from dynamic ones (such as expression and gaze; Haxby, Hoffman, & Gobbini 2000). In fact, our findings present the best evidence yet of such a division of labor between identifiable nodes in the macaque brain, opening the door to further characterization of putative static and dynamic streams by electrophysiological and causal approaches. This could ultimately elucidate how the myriad signals conveyed by faces are given meaning by the brain (Calder & A. W. Young 2005; Barraclough & Perrett 2011) at neuron and network levels. In this way, the specializations for facial motion within the areas described here provide a concrete anatomical framework for investigating both the computations that extract and abstract from facial dynamics and, more generally, the interrelated neural representations of form and motion.

Responses to the Conjunction of Head and Body

When we observe a face atop a body, these two distinct forms fundamentally represent the same agent with the same identity, feelings, and motivations. This is why poker players scrutinize both faces and bodies for tells, why it is so alarming to recognize a friend from behind but find yourself face to face with a stranger when he turns around, and why, presumably, the brain is wired to use the body as well as the face to interpret emotion and identity (Aviezer, Trope, & Todorov 2012; Rice et al. 2013). Our investigation into the

effects of viewing faces in the context of bodies aimed to identify regions of the macaque temporal lobe that jointly process these two related forms. Our work shows that activity within some face patches is elicited specifically by the joint observation of head and body, contrary to the view that the representations of faces and bodies are separated into independent domains in the temporal lobe.

When comparing responses to whole monkeys with the sum of responses to heads and bodies, we found 3 different activity patterns in the face and body patches: sublinear responses in PL, ML and the body patches, linear responses in MF and MD, and a supralinear response in AF and possibly AL (Figure 11B). The sublinear and linear responses are in line with the common view of face and body patches as being largely non-selective for forms that are not their preferred stimulus. Both monkey electrophysiology (e.g. Sato 1989; Miller, Gochin, & Gross 1993; Zoccolan, Cox, & DiCarlo 2005) and human neuroimaging (e.g. Beck & Kastner 2005; Gentile & Jansma 2010; Nagy, Greenlee, & Kovács 2011) have shown that viewing multiple stimuli together can lead to less activation in high-level visual areas than would be naively predicted by a linear model based on responses to independently-presented stimuli. This is usually understood as the result of competition between different objects for a limited pool of neural resources (reviewed in Desimone & Duncan 1995). In this light, the sublinear responses in PL and ML can be explained by the competition between the preferred face stimuli and the non-preferred body stimuli. The same logic holds true for the body patches, though in this case the body stimuli would be the preferred subjects. Linearity would be predicted, on their other hand, if the signal within a measured region were composed of face and body selective components that do not interact even when a face is presented with a body. An

alternative explanation for linearity is that a body attached to a face may be treated as part of the face. Because the combination of body and face would form a single perceptual group, no competition would be induced (McMains & Kastner 2010; Bernstein et al. 2014). If this is the case, the progression from sublinear to linear when moving from PL and ML to MF and MD could represent a transition to a unified representation of faces and bodies (as argued in Bernstein et al. 2014).

In contrast to these more readily explainable responses, the supralinear response that characterizes AF and AL (at least when form-specificity is accounted for), and which we sometimes found in whole brain analyses, came as a surprise. Supralinear fMRI signals have not normally been evoked by viewing multiple objects at once (MacEvoy & Epstein 2009; Nagy, Greenlee, & Kovács 2011; Baeck, Wagemans, & Op de Beeck 2013), so their presence here suggests a special link between faces and bodies in these patches. In support of such a link, the supralinear responses in AF and AL depended on the specific form of the body; faces shown atop non-body objects in these patches elicited merely linear responses (Figure 12B). In contrast to the form-selectivity of the supralinear responses in AL and AF, the sublinear and linear responses to faces with bodies in the other patches mirrored the response patterns evoked when faces were presented in conjunction with non-body objects. This supports the idea that the increased activity in anterior patches AL and AF is fundamentally different than the response to bodies in more posterior face patches, and results from the unique relationship between faces and bodies in these regions.

Interestingly, the antSTS body patch also demonstrated a form-dependent interaction effect, showing a greater sublinearity when a face was paired with a

body than when it was paired with a non-body object (Figure 12C). The form-specific suppression effect speaks to a possible functional link between the representations of faces and bodies in this patch. The fact that no similar effect was found in midSTS further bolsters the idea that there is an increasing integration of face and body information in the more anterior parts of the temporal lobe. This may be one facet of the frequently noted increase in receptive field size and development of complex stimulus selectivity as information moves from posterior to anterior through the visual stream (Rust & DiCarlo 2010).

While there are a number of challenges to comparing our work here with human research, human studies have also pointed to an increasing integration of face and body information in more anterior areas. The three primary barriers to comparison with the existing human studies are the diversity of approaches that these studies took, the fMRI signatures that they identified, and the small number of face areas that they examined. Only three human studies have shared our basic approach, employing univariate analysis to explore the relative responses to faces, bodies, and whole agents in multiple areas (Morris, Pelphrey, & McCarthy 2006; Song et al. 2013; Kaiser et al. 2014), and no two of these defined their regions of interest in the same way. Song et al. (2013) – like us – chose to examine activity within independently localized face and body areas, but they – like other research groups working with humans – found no areas with supralinear responses to faces and bodies viewed together. One explanation for this lack of a supralinear signature is that human experiments have only examined functional regions at the posterior end of the temporal lobe, while the supralinear responses that we found were only evident in very anterior face patches. Alternatively, this disparity might result from our use of an fMRI

contrast agent, while all previous human studies have measured functional activity using a functionally similar, but not equivalent, endogenous fMRI signal (see discussion above).

Despite these differences, all human studies that have examined multiple face areas in the brain have found a more integrated representation of face and body in the anterior (FFA) than in the posterior (OFA), much as we did. One study showed that altering a body could reverse fMRI adaptation in the FFA but not OFA (Andrews et al. 2010), a second study showed more supralinear fMRI adaptation to faces and bodies in FFA than OFA (Schmalzl, Zopf, & Williams 2012), another found an increased response when a face was presented with a body only in FFA and not OFA (Song et al. 2013), and the fourth suggested competition between faces and their bodies in OFA but not FFA (Bernstein et al. 2014). Thus, while the human and macaque face processing systems cannot yet be homologized area to area, the overall organization of body context effects within them is similar.

Human studies also inspired our investigation into the role that facial ambiguity plays in the joint responses to faces and bodies. We found a supralinear effect of providing body context to blurred faces in AF and a suggestion of it in AL, the same patches where positive body context effects were seen with unambiguous faces (Figure 13B, top). It is remarkable that there is a supralinear response to bodies and heads in these face patches even in the total absence of a clearly defined face. This is reminiscent of the psychological finding that a faceless head paired with a body, when viewed briefly, is often perceived as having a face and even shows inversion effects like a face (Brandman & Yovel 2012), suggesting that face-processing mechanisms can be engaged by bodies

paired with the general shape of a head . However, our data argues against the related idea that the activity induced by blurred faces atop bodies represents the normal recruitment of the face-processing machinery usually activated by unambiguous faces. In contrast to what Cox et al. (2004) found in the human FFA, the contextually evoked activity elicited by ambiguous faces atop bodies was equal to the contextual activity elicited by unambiguous faces atop bodies (Figure 13B, bottom). Perhaps the contextual activity evoked by a body paired with either an ambiguous or unambiguous face is a whole-agent response that is not dependent upon internal features of the face. Such a whole-agent response could either originate from a collection of otherwise silent neurons in the face patches, or, alternatively, from face neurons firing beyond their usual response to face features. Direct recordings of neurons in AL and AF would be the most straightforward way to explore these possibilities, and could also confirm that the supralinear fMRI signals we observed accurately reflects a supralinear change in neural firing.

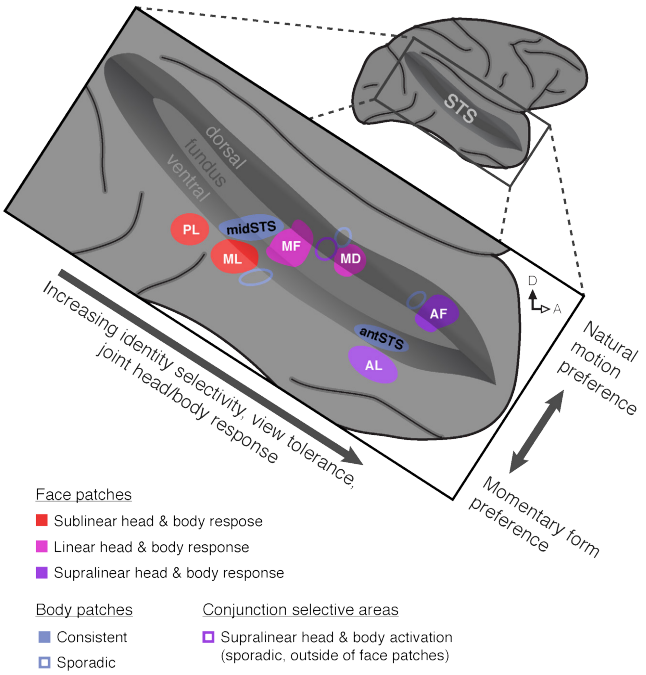
While face-selective areas provide a useful framework for exploring the joint representation of faces and bodies in the temporal lobe, our findings also suggest that focusing our attention solely on these regions may not tell the whole story. In the two subjects where a whole-brain analysis was sufficiently powered to visualize positive interactions between face and body responses, we found supralinear responses not just in the anterior STS near face patches AL and AF, but also in the mid-STS near face patches MF and MD (Figure 10, M1 and M4). This is in seeming contradiction with the fact that neither of these middle face patches showed a supralinear response in our group analysis, but agrees with the only previous macaque study that described a specific responsiveness to whole

agents in the temporal lobe. Wachsmuth, Oram, and Perret (1994) found neurons with selectivity for whole people (but not faces or bodies) in the upper bank and fundus of the STS, and the coordinates of their recording sites suggest that these cells fell near the middle face patches. The fact that these supralinear responses in the middle STS were evident in our whole brain analysis but not the group analysis of MD and MF suggests that whole-agent-responsive regions may make up only part of these face patches, or may reside in neighboring but distinct functional areas. This finding also calls into question whether the supralinear responses in AL and AF originate from the previously characterized face-selective cells within them (Freiwald & Tsao 2010; McMahon et al. 2014), or instead from a unrecognized population of cells selective for faces with bodies in an overlapping or closely neighboring functional area. To distinguish these possibilities, it will be necessary to perform a study with enough power to reliably localize areas with a supralinear response to faces and bodies and then use independent data to explore if these areas have selectivity for faces presented alone. Even then, recordings of single neurons will be needed to definitively determine whether selectivity for lone faces and for faces with bodies can be found within the same cell.

Our examination of the joint representation of heads and bodies expands our understanding of biological form processing in the temporal lobe, revealing an intertwined functional organization that is richer than previously appreciated (Figure 15). Within the recognized face patches, the temporal lobe contains regions that select for the conjunction of head and body. While there is limited evidence for independent areas of joint form selectivity at this point, face patches along the STS form a posterior to anterior hierarchy that demonstrates the

Figure 15. Model of Face and Body Representations Along the Macaque Temporal Lobe

Viewing faces or bodies activates a number of functional areas along the STS. Near the face patches, two large and reliable body selective patches can be found in every hemisphere (midSTS and antSTS; solid blue). Additional regions of body selectivity can be found in some hemispheres anterior to ML, near MD, and near AF (blue outlines). Moving posterior to anterior, different face patches show different responses to the simultaneous sight of faces and bodies. The most posterior patches (PL and ML; red) show a sublinear response to faces and bodies, patches in the middle (MF and MD; magenta) show a nearly linear response, and the most anterior (AF and AL; purple) show a likely supralinear response. Notably, only the supralinear responses are body-specific (Figure 12). The body patches show sublinear responses to faces and bodies presented together. In addition, whole brain analysis suggests that there can sometimes be areas of supralinear response to a face atop a body in the STS near MF and MD, and such supralinear areas may fall outside of face patches (purple outline).


STS
dorsal
fundus
ventral
PL
midSTS
ML
MF
MD
AF
antSTS
AL
D
A
Increasing identity selectivity, view tolerance, joint head/body response
Natural motion preference
Momentary form preference
Face patches
Sublinear head & body respose
Linear head & body response
Supralinear head & body response
Body patches
Consistent
Sporadic
Conjunction selective areas
Supralinear head & body activation (sporadic, outside of face patches)

emergence of this positive interaction between head and body, with posterior patches showing sublinear responses to the conjunction of these forms, middle patches showing linear responses, and anterior patches showing supralinear responses. The ability to examine the differences in the interaction of head and body representations between these identifiable regions has the potential to shed light on major problems of object vision: how multiple objects are simultaneously encoded at different stages in the temporal lobe (Zoccolan, Cox, & DiCarlo 2005), how selectivity for complex objects is developed (Connor, Brincat, & Pasupathy 2007; Rust & DiCarlo 2010), and how spatial context facilitates object processing (Bar 2004). The functional specializations identified in this set of experiments provide a framework for understanding the neural mechanisms that integrate visual information from faces and bodies, and may shed light on the more general question of how the brain combines information from spatially disparate but meaningfully intertwined parts of the visual world.

How Contextual Effects Illuminate the Organization of the Face Patches

In acknowledging visual context's importance, we have attempted to move beyond the limitations of motionless, isolated face stimuli to study two very different contextual effects within the macaque face processing system. While each study alone enriched our understanding of the considered contextual cue and the organization of the face patches, combining the findings of the two studies both strengthens these gains and provides an even deeper understanding of face processing in the macaque.

At a very practical level, both of these experiments revealed that the reciprocal exchange of activity between the face patches is not strong enough to

obscure local differences in specialization. The face patches form a densely interconnected web of functional areas (Moeller, Freiwald, & Tsao 2008), and previous univariate fMRI experiments have revealed response profiles of the different patches that differ in degree, but not in fundamental character (e.g. Polosecki et al. 2013; Russ & Leopold 2015). Thus, it had seemed possible that activity originating in a single patch might spread throughout the network more quickly than fMRI could capture, and that fMRI might therefore be useless for functionally separating different nodes within this system. The fact that both of our experiments revealed patterns of fMRI activity that are localized, and not shared throughout the entire face-processing system, precludes this possibility.

Fundamentally, these localized differences in visible activity mean that fMRI can be used to define functional signatures specific to a face patch, and researchers can use these to identify equivalent functional areas across animals. For instance, AF was the only face patch in our studies to show both selectivity for natural motion and a supralinear response to co-presented faces and bodies; this signature can now be used to recognize a functionally homologous area in other monkeys, even if unusual anatomical patterning otherwise obscures AF. Conceptually, this finding also suggests that fMRI activity patterns represent the neural responses within each individual face patch, rather than composite activity of the entire face-processing system. This supports the utility of fMRI as a tool for targeting high-resolution experimental methods to functionally interesting regions of the face processing system.

At the level of functional neuroanatomy, these studies also show that the face patches are not solely driven by face form, nor are they organized in a wholly different manner than the rest of the temporal lobe visual areas. Both

motion and body form drive activity in some patches, and these effects are not uniform throughout the system, but fall along axes of stimulus selectivity that have been previously suggested by years of human and monkey research. The dorsal-ventral split in sensitivity to natural face motion is in line with the influential idea that vision is processed through a ventral "what" stream and a dorsal "how" stream (Goodale & Milner 1992), as well as the suggestion that the dorsal bank of the macaque STS is a point of communication between these streams (Oram & Perrett 1996). The posterior-to-anterior increase in the joint response to faces and bodies likewise parallels the increased complexity of preferred features demonstrated by neurons at the anterior end of the temporal lobe (Connor, Brincat, & Pasupathy 2007; Rust & DiCarlo 2010). Considered as a pair, these trends highlight the extent to which the face patches appear to be representative of the wider visual system, even as they are functionally and anatomically separate from it.

Considering the intersection of these two experiments can also suggest alternative functional organization schemes. Although we considered facial motion and body context separately, the representations of these two features may be related in certain cases. The strong effects of both temporal and spatial context in face patch AF, for instance, could point to a higher-level specialization within this area. Previous electrophysiological studies in the anterior STS near AF have revealed a wide range of specializations for biological perception – ranging from neurons that respond to paired sound and motion to neurons that respond to the presence of a visually obscured person – leading some to speculate that this region is a hub of social cognition (reviewed in Brothers 1996; Keysers & Perrett 2004). Understanding both facial movements and body posture

is critical for social interactions, and the fact that AF is attuned these influences raises the possibility that it may be part of, or even encompass, the hypothesized social cognition area. Further characterization of the face patches could explore whether AF plays a role in additional, untested, social functions.

Finally, and perhaps most critically, both of these studies show that contextual cues can exert their effects within the visual processing stream, and not solely within the "higher", more cognitive areas of the brain. By identifying areas of the macaque cortex where this happens, these experiments open the door to experiments that can probe the cellular and system-level origins of this contextual influence.

Next Steps

Using macaque fMRI to study how the brain processes facial context has allowed us to rely upon information from both human fMRI experiments and macaque electrophysiology, along with the structure provided by the face patch system. By doing so, we have identified temporal lobe areas that react to the context in which a face is viewed and have hypothesized both the underlying neural activity in these areas and their behavioral impact. However, in order to understand the processing of facial context from neuron to human behavior using the macaque as a model, three additional steps need to be taken. First, the contextual effects we observed must be demonstrated at the level of single neurons. Second, a causative link must be established between the activity of these neurons and the ability to act on the contextual information. Finally, there needs to be a better understanding of how (and if) the organization of the

macaque face patches can be mapped onto human face areas. Our experiments have laid the groundwork for studies that can meet these needs.

Although the wide field of view provided by fMRI makes it a useful complement to single-cell observation, it can only hint at the richness of the information carried by neurons within each voxel (see Dubois, de Berker, & Tsao 2015 for a pertinent example). While new technologies are expanding the ways that single neurons can be monitored, the simplest method for understanding the activity underlying the fMRI signatures we observed remains electrophysiological recording. In the short time since the macaque face patches were discovered, electrophysiological studies have already begun to piece together the tuning of face patch neurons, revealing their dependence on face features (Freiwald, Tsao, & Livingstone 2009; Issa & DiCarlo 2012), their patch-dependent selectivity for identity and viewpoint (Freiwald & Tsao 2010), and the stability of their tuning over multiple months (McMahon et al. 2014). Our experiments have suggested recording sites where neurons representing facial motion and the interaction of face and body form can be found, recorded, and characterized within the face processing system.

Comparing the activity of neurons within MD and AF (which we predict respond to natural motion) to the activity of neurons in PL, ML, and AL (which we predict respond to momentary static form) would be the logical place to begin an electrophysiological exploration of how cells represent facial motion. While the first step should be to see if single neurons recapitulate the patterns of natural motion preference we see in fMRI, electrophysiological experiments would also be an effective way to define what type of motion excites these face

cells, over what time window they integrate motion, and how these cells combine information about dynamics and form.

The face videos that we used in our first study contained many types of motion, including head movements, chewing, and expressive face motion. By restricting presented stimuli to a specific rigid motion (such as translation, looming, or rotation in or through the visual plane) or non-rigid motion (such as a mouth, eye, or brow movement), it would be possible to determine what types of movements are represented within the identified face patches. To explore how motion is integrated over time, stimuli in which a motion is reversed or scrambled at different time scales (see Hasson et al. 2008; Singer & Sheinberg 2010) could be utilized to determine what temporal aspects of an effective motion stimulus are necessary to elicit a neural response. Animated avatars, either made with traditional 3D modeling (as in Hill & A. Johnston 2001) or derived from videos of animals (as in Nagle et al. 2013), could be used to show different monkeys performing identical actions; this would allow the definition of joint tuning functions for form and dynamics for neurons in the ventral and dorsal portions of the face processing system. While these three suggested experiments would only scratch the surface of possible selectivities a cell tuned for facial motion might exhibit, they could suggest what behavioral functions the dorsal face patches support.

Similarly, recordings in AF and AL could both confirm that individual face cells can integrate contextual body information and determine the way in which they integrate it. Experiments that we recounted earlier suggest that both the relative location of a face and body, as well as the specific form of a body, can alter the contextual effect that a body exerts. By translating, rotating, or flipping a

presented body while holding a face constant, it would be possible to investigate the location dependency of the neural body context effects. Exhaustively pairing a set of faces with a set of bodies could provide insight into the nature of these face/body interactions. For instance, a given body might have an equivalent contextual effect on a neuron's firing no matter the face with which it is paired. Alternatively, the strength of contextual effects could depend on the congruency of affect between face and body, the rotational position of the head relative to the body, or other such interactions. In the event that it is possible to confirm the existence of areas with a supralinear response to faces and bodies *outside* of the face patches, single cell recordings would also be critical for confirming whether such areas contain a high density of neurons specifically selective for whole agents. These studies would suggest whether the areas we have identified underlie the known psychological links between body form and face form.

While electrophysiology could reveal the variety of contextual signals contained within the face patches, neural inactivation will be necessary to understand whether these signals are necessary for acting upon contextual information. The GABA agonist muscimol and optogenetic inhibition techniques have already been used in and near the face patches to show the role of these areas in involuntary gaze-following (Roy, Shepherd, & Platt 2012) and facial discrimination (Afraz, Boyden, & DiCarlo 2014). By selectively inactivating the face patches thought to be involved in contextual information processing, it will be possible to determine whether the observed contextual activity signatures underlie the expected behavior.

For instance, our results suggest that inactivating MD and AF would reduce the ability to discriminate identities embodied by dot-motion representations of

faces, but not by still photographs of faces. Through similar logic, inactivation of AF and AL would be expected to decrease the influence of body posture on reported facial emotion. The most difficult part of performing such experiments would likely not be implementing the inactivation protocol, but designing tasks that can measure such complicated concepts as perceived facial emotion and training monkeys to reliably perform them. Such tasks will ultimately be necessary, however, to understand if and how macaques use contextual information.

As an aside, inactivation techniques could also help to reveal the connectivity underlying contextual activity. The discovery that activity within the face patches can be modulated by the presence of a body raises the possibility that there may be direct interactions between neighboring face and body patches. By inactivating body patches and seeing if supralinear body context responses remain in the anterior face patches, it would be possible to determine whether the joint representation of face and body that we observed is dependent upon prior activity within the body patches, or if the form-selectivity of body context effects arises from an independent mechanism. A similar method has already been used to demonstrate that the face patches are hierarchically organized, showing that pharmacologic inactivation of the middle face patches blocks the activity of anterior face patches (Liu et al. 2011). If the activity of body patches is not required for the context effects that we observed, it may even be possible to disrupt the normal perception of bodies while leaving their effects on face perception intact.

Even if context responses were understood at the level of single cells, and inactivation of these cells were to elicit the expected behavioral deficit, the ability

to apply these findings to the human brain would be hindered by the current lack of known homology between macaque and human face areas. While our experiments have uncovered fMRI signatures which can be used to identify specific face patches in macaques, at the moment these signatures cannot be identified within the human brain; human experiments have uncovered neither face areas that respond more to scrambled motion than to natural motion (Schultz et al. 2013) nor face areas with a supralinear response to faces paired with bodies (Kanwisher, Stanley, & Harris 1999; Morris, Pelphrey, & McCarthy 2006; Song et al. 2013; Kaiser et al. 2014). However, by performing human fMRI experiments that measure CBV rather than BOLD signal, it might be possible to identify human face areas that behave much like the macaque face patches do.

As we mentioned earlier, BOLD signal and cerebral blood volume diverge in certain cases. For instance, CBV changes can sometimes be found in areas with no BOLD response (Smirnakis et al. 2007) and while a positive BOLD signal is generally accompanied by an increase in CBV, a negative BOLD signal may *also* be accompanied by an increase in CBV (Goense, Merkle, & Logothetis 2012). These differences between BOLD and CBV responses could explain the continued difficulty in aligning the functional profiles of macaque and human face areas.

By measuring human CBV responses, it might therefore be possible to find face areas with functional signatures similar to those that we found in macaques. Researchers have recently begun to measure CBV in humans with Feraheme, one of the contrast agents we used (Qiu et al. 2012; D'Arceuil et al. 2013). However, given the discovery that Feraheme or its metabolites can accumulate and hinder future brain imaging (Lafer-Sousa & Conway 2013; Gagin et al. 2014), it may not

be appropriate to use exogenous contrast to measure human CBV until a superior contrast agent is developed. An alternative way to measure human CBV would be to employ vascular-space-occupancy (VASO) imaging, an MRI method that requires no exogenous contrast agent and relies upon a specialized MRI pulse sequence to determine the volume of blood within each voxel (Lu et al. 2003; reviewed in Lu, Hua, & Zijl 2013). Although VASO imaging has worse functional sensitivity than BOLD imaging, the additional time required to measure a meaningful signal would likely be worth the ability to directly compare the resulting data to the CBV data from macaque experiments. This comparison could be a key step in understanding the neuron-level origins of a complex perceptual ability.

No part of our everyday world exists in a contextual vacuum. Faces are no exception to this rule, and, taking advantage of this fact, we have identified regions of form-selective visual cortex that reflect contextual influences. While future studies – like those just described – will be necessary to understand the role that these areas play in creating comprehensible visual experience, our findings reinforce the importance of recognizing the linked challenges and opportunities that visual context presents. Embracing these challenges and opportunities might ultimately provide the perspective we need to perceive the whole of the brain's form-processing machinery, and not just its constituent parts.

References

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition*, *52*(1), 61–69.

Afraz, A., Boyden, E. S., & DiCarlo, J. J. (2014, August 22). Optogenetic and pharmacological suppression of face-selective neurons reveal their causal role in face discrimination behavior. *Vision Sciences Society 2014*. Association for Research in Vision and Ophthalmology. doi:10.1167/14.10.600

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*(5), 403–410. doi:10.1111/j.0956-7976.2005.01548.x

Andrews, T. J., Davies-Thompson, J., Kingstone, A., & Young, A. W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience*, *30*(9), 3544–3552. doi:10.1523/JNEUROSCI.4863-09.2010

Aparicio, P. (2014, February 11). Functional and structural characterization of the macaque middle face patch. (J. J. DiCarlo, Ed.). Cambridge, MA: Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/87456

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion (Washington, D.C.)*, *11*(6), 1406–1414. doi:10.1037/a0023578

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., et al. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*(7), 724–732. doi:10.1111/j.1467-9280.2008.02148.x

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229. doi:10.1126/science.1224313

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *NeuroImage*, *70*, 37–47. doi:10.1016/j.neuroimage.2012.12.023

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. doi:10.1038/nrn1476

Barraclough, N. E., & Perrett, D. I. (2011). From single cells to social perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1571), 1739–1752. doi:10.1073/pnas.171288598

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(3), 373–379.

Bassili, J. N. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*(11), 2049–2058. doi:10.1002/bs.3830070216/abstract

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*(8), 1110–1116. doi:10.1038/nn1501

Bell, A. H., Malecek, N. J., Morin, E. L., Hadj-Bouziane, F., Tootell, R. B. H., & Ungerleider, L. G. (2011). Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *Journal of Neuroscience*, *31*(34), 12229–12240. doi:10.1523/JNEUROSCI.5865-10.2011

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, *57*(1), 289–300.

Bernstein, M., Oron, J., Sadeh, B., & Yovel, G. (2014). An integrated face–body representation in the fusiform gyrus but not the lateral occipital cortex. *Journal of Cognitive Neuroscience*, *26*(11), 2469–2478. doi:10.1162/jocn_a_00639

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *8*(6), 510–523.

Brandman, T., & Yovel, G. (2012). A face inversion effect without a face. *Cognition*, *125*(3), 365–372. doi:10.1016/j.cognition.2012.08.001

Brothers, L. (1996). Brain mechanisms of social cognition. *Journal of Psychopharmacology (Oxford, England)*, *10*(1), 2–8. doi:10.1177/026988119601000102

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology (London, England : 1953)*, *77 (Pt 3)*, 305–327.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651. doi:10.1038/nrn1724

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. doi:10.1017/S0140525X12000477

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*(2), 140–147. doi:10.1016/j.conb.2007.03.002

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science*, *304*(5667), 115–117. doi:10.1126/science.1093110

Cunningham, D. W., & Wallraven, C. (2009). The interaction between motion and form in expression recognition (pp. 41–44). Presented at the ACM Symposium on Applied Perception in Graphics and Visualization 2009, New York, USA: ACM. doi:10.1145/1620993.1621002

D'Arceuil, H., Coimbra, A., Triano, P., Dougherty, M., Mello, J., Moseley, M., et al. (2013). Ferumoxytol enhanced resting state fMRI and relative cerebral blood volume mapping in normal human brain. *NeuroImage*, *83*, 200–209. doi:10.1016/j.neuroimage.2013.06.066

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. doi:10.1146/annurev.ne.18.030195.001205

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

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. doi:10.1016/j.neuron.2012.01.010

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. doi:10.1126/science.1063414

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*(6), 2791–2802. doi:10.1523/JNEUROSCI.4037-14.2015

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. doi:10.1016/j.neuroimage.2007.09.034

Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Statistical Science*, *19*(4), 676–685.

Farivar, R., & Vanduffel, W. (2014). Functional MRI of awake behaving macaques using standard equipment. In T. D. Papageorgiou, G. I. Christopoulos, & S. M. Smirnakis (Eds.), *Advanced Brain Neuroimaging Topics in Health and Disease - Methods and Applications*. doi:10.5772/31413

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), 17–17. doi:10.1167/11.3.17

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. doi:10.1016/j.neuroimage.2012.01.021

Fisher, C., & Freiwald, W. A. (2015). Contrasting specializations for facial motion within the macaque face-processing system. *Current Biology*, *25*(2), 261–266. doi:10.1016/j.cub.2014.11.038

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651. doi:10.1002/hbm.20630

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*(6005), 845–851. doi:10.1126/science.1194908

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*(9), 1187–1196. doi:10.1038/nn.2363

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. doi:10.1098/rstb.2005.1622

Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *NeuroImage*, *10*(1), 1–5. doi:10.1006/nimg.1999.0439

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., & Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *Journal of Neuroscience*, *32*(45), 15952–15962. doi:10.1523/JNEUROSCI.1992-12.2012

Gagin, G., Bohon, K. S., Connelly, J., & Conway, B. R. (2014, November 19). fMRI signal dropout in rhesus macaque monkey due to chronic contrast agent administration. *Neuroscience 2014*. Washington, DC.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504.

Gentile, F., & Jansma, B. M. (2010). Neural competition through visual similarity in face selection. *Brain Research*, *1351*, 172–184. doi:10.1016/j.brainres.2010.06.050

Gerits, A., & Vanduffel, W. (2013). Optogenetics in primates: a shining future? *Trends in Genetics*, *29*(7), 403–411. doi:10.1016/j.tig.2013.03.004

Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Current Biology*, *20*(1), 32–36. doi:10.1016/j.cub.2009.10.077

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–192. doi:10.1038/nrn1057

Goense, J. B. M., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, *18*(9), 631–640. doi:10.1016/j.cub.2008.03.054

Goense, J. B. M., Merkle, H., & Logothetis, N. K. (2012). High-resolution fMRI reveals laminar differences in neurovascular coupling between positive and negative BOLD responses. *Neuron*, *76*(3), 629–639. doi:10.1016/j.neuron.2012.09.019

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

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

Grimaldi, P. C., Saleem, K. S., & Tsao, D. Y. (2012, October 14). Anatomical connections of functionally defined anterior face patches in the macaque monkey. *Neuroscience 2012*. New Orleans, LA.

Grimaldi, P. C., Saleem, K. S., & Tsao, D. Y. (2013, November 13). Subcortical connections of the functionally-defined face patches in the macaque monkey. *Neuroscience 2013*. San Diego, CA.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(7), 1564–1576. doi:10.1093/cercor/bhr226

Hadj-Bouziane, F., Bell, A. H., Knusten, T. A., Ungerleider, L. G., & Tootell, R. B. H. (2008). Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5591–5596. doi:10.1073/pnas.0800489105

Handwerker, D. A., Gonzalez-Castillo, J., D'Esposito, M., & Bandettini, P. A. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage*, *62*(2), 1017–1023. doi:10.1016/j.neuroimage.2012.02.015

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: facial expressions of emotions, in context. *Emotion Review*, *5*(1), 60–65. doi:10.1177/1754073912451331

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. doi:10.1523/JNEUROSCI.5487-07.2008

Haxby, J. V., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it? *Journal of Cognitive Neuroscience*, *20*(12), 2125–2136. doi:10.1162/jocn.2008.20148

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

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. doi:10.1038/71152

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. doi:10.2307/4615733

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, *34*(4), 247–261.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia*, *31*(2), 173–181.

Issa, E. B., & DiCarlo, J. J. (2012). Precedence of the eye region in neural processing of faces. *Journal of Neuroscience*, *32*(47), 16666–16682. doi:10.1523/JNEUROSCI.2391-12.2012

Issa, E. B., Papanastassiou, A. M., & DiCarlo, J. J. (2013). Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *Journal of Neuroscience*, *33*(38), 15207–15219. doi:10.1523/JNEUROSCI.1248-13.2013

Janssens, T., Keil, B., Farivar, R., McNab, J. A., Polimeni, J. R., Gerits, A., et al. (2012). An implanted 8-channel array coil for high-resolution macaque MRI at 3T. *NeuroImage*, *62*(3), 1529–1536. doi:10.1016/j.neuroimage.2012.05.028

Janssens, T., Zhu, Q., Popivanov, I. D., & Vanduffel, W. (2014). Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *Journal of Neuroscience*, *34*(31), 10156–10167. doi:10.1523/JNEUROSCI.2914-13.2013

Jastorff, J., Popivanov, I. D., Vogels, R., Vanduffel, W., & Orban, G. A. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *NeuroImage*, *60*(2), 911–921. doi:10.1016/j.neuroimage.2011.12.087

Jellema, T., & Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*, *41*(13), 1728–1737. doi:10.1016/S0028-3932(03)00175-1

Johnston, P., Mayes, A., Hughes, M., & Young, A. W. (2013). Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(9), 2462–2472. doi:10.1016/j.cortex.2013.01.002

Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *Journal of Neurophysiology*, *111*(1), 82–90. doi:10.1152/jn.00371.2013

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

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport*, *10*(1), 183–187.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences*, *8*(11), 501–507. doi:10.1016/j.tics.2004.09.005

Kim, D.-S., Ronen, I., Olman, C., Kim, S. G., Ugurbil, K., & Toth, L. J. (2004). Spatial relationship between neuronal activity and BOLD functional MRI. *NeuroImage*, *21*(3), 876–885. doi:10.1016/j.neuroimage.2003.10.018

Kim, S. G., & Ogawa, S. (2012). Biophysical and physiological origins of blood oxygenation level-dependent fMRI signals. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *32*(7), 1188–1206. doi:10.1038/jcbfm.2012.23

Knight, B., & Johnston, A. (1997). The Role of Movement in Face Recognition. *Visual Cognition*, *4*(3), 265–273. doi:10.1080/713756764

Kolster, H., Mandeville, J. B., Arsenault, J. T., Ekstrom, L. B., Wald, L. L., & Vanduffel, W. (2009). Visual field map clusters in macaque extrastriate visual cortex. *Journal of Neuroscience*, *29*(21), 7031–7039. doi:10.1523/JNEUROSCI.0518-09.2009

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emotion Review*, *5*(1), 41–46. doi:10.1177/1754073912451349

Ku, S.-P., Tolias, A. S., Logothetis, N. K., & Goense, J. B. M. (2011). fMRI of the Face-Processing Network in the Ventral Temporal Lobe of Awake and Anesthetized Macaques. *Neuron*, *70*(2), 352–362. doi:10.1016/j.neuron.2011.02.048

Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature Neuroscience*, *16*(12), 1870–1878. doi:10.1038/nn.3555

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, *27*(6), 974–985. doi:10.3758/BF03201228

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(3), 567–576. doi:10.1093/cercor/bhr119

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., et al. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, *465*(7299), 788–792. doi:10.1038/nature09108

Leibovici, D. G., & Smith, S. (2000). *Comparing groups of subjects in fMRI studies: a review of the GLM approach* (No. TR00DL1). Oxford, UK. Retrieved from http://www.fmrib.ox.ac.uk/analysis/techrep/tr00dl1/tr00dl1/index.html

Lemay, G., Kirouac, G., & Lacouture, Y. (1995). Expressions faciales émotionnelles spontanées dynamiques et statiques: comparaison d'études de jugement catégoriel et dimensionnel. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *27*(2), 125–139. doi:10.1037/0008-400X.27.2.125

Liu, N., Jones, K. B., Hadj-Bouziane, F., Turchi, J. N., Tootell, R. B. H., & Ungerleider, L. G. (2011, November 14). Hierarchical organization of face-selective regions in macaque cortex as revealed by fMRI and pharmacological deactivation. *Neuroscience 2011*. Washington, DC.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157. doi:10.1038/35084005

Longmore, C. A., & Tree, J. J. (2013). Motion as a cue to face recognition: evidence from congenital prosopagnosia. *Neuropsychologia*, *51*(5), 864–875. doi:10.1016/j.neuropsychologia.2013.01.022

Lu, H., Golay, X., Pekar, J. J., & van Zijl, P. C. M. (2003). Functional magnetic resonance imaging based on changes in vascular space occupancy. *Magnetic Resonance in Medicine : Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *50*(2), 263–274. doi:10.1002/mrm.10519

Lu, H., Hua, J., & Zijl, P. C. M. (2013). Noninvasive functional imaging of cerebral blood volume with vascular-space-occupancy (VASO) MRI. *NMR in Biomedicine*, *26*(8), 932–948. doi:10.1002/nbm.2905

MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, *19*(11), 943–947. doi:10.1016/j.cub.2009.04.020

Mandeville, J. B., Choi, J.-K., Jarraya, B., Rosen, B. R., Jenkins, B. G., & Vanduffel, W. (2011). fMRI of cocaine self-administration in macaques reveals functional inhibition of basal ganglia. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(6), 1187–1198. doi:10.1038/npp.2011.1

Mandeville, J. B., Marota, J. J., Kosofsky, B. E., Keltner, J. R., Weissleder, R., Rosen, B. R., & Weisskoff, R. M. (1998). Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magnetic Resonance in Medicine : Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *39*(4), 615–624.

McDonald, S., & Saunders, J. C. (2005). Differential impairment in recognition of emotion across different media in people with severe traumatic brain injury. *Journal of the International Neuropsychological Society : JINS*, *11*(4), 392–399. doi:10.1017/S1355617705050447

McMahon, D. B. T., Jones, A. P., Bondar, I. V., & Leopold, D. A. (2014). Face-selective neurons maintain consistent visual responses across months. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8251–8256. doi:10.1073/pnas.1318331111

McMains, S. A., & Kastner, S. (2010). Defining the units of competition: influences of perceptual organization on competitive interactions in human visual cortex. *Journal of Cognitive Neuroscience*, *22*(11), 2417–2426. doi:10.1162/jocn.2009.21391

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16518–16523. doi:10.1073/pnas.0507650102

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. doi:10.1073/pnas.1112895108

Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research*, *616*(1-2), 25–29.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417. doi:10.1016/0166-2236(83)90190-X

Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science*, *320*(5881), 1355–1359. doi:10.1126/science.1157436

Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2006). Occipitotemporal activation evoked by the perception of human bodies is modulated by the presence or absence of the face. *Neuropsychologia*, *44*(10), 1919–1927. doi:10.1016/j.neuropsychologia.2006.01.035

Nagle, F., Griffin, H., Johnston, A., & McOwan, P. (2013). Techniques for mimicry and identity blending using morph space PCA. In *Computer Vision - ACCV 2012 Workshops* (Vol. 7729, pp. 296–307). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-642-37484-5_25

Nagy, K., Greenlee, M. W., & Kovács, G. (2011). Sensory competition in the face processing areas of the human brain. *PLoS ONE*, *6*(9), e24450. doi:10.1371/journal.pone.0024450

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., et al. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, *31*(10), 3743–3756. doi:10.1523/JNEUROSCI.4803-10.2011

Nelissen, K., Vanduffel, W., & Orban, G. A. (2006). Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. *Journal of Neuroscience*, *26*(22), 5929–5947. doi:10.1523/JNEUROSCI.0824-06.2006

Ohayon, S., Freiwald, W. A., & Tsao, D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron*, *74*(3), 567–581. doi:10.1016/j.neuron.2012.03.024

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, *76*(1), 109–129.

Orban, G. A., Zhu, Q., & Vanduffel, W. (2014). The transition in the ventral stream from feature to real-world entity representations. *Frontiers in Psychology*, *5*, 695. doi:10.3389/fpsyg.2014.00695

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*(1), 603–608. doi:10.1152/jn.00513.2004

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., et al. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *The Journal of Experimental Biology*, *146*, 87–113.

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

Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C., & Oram, M. W. (2009). Seeing the future: natural image sequences produce "anticipatory" neuronal activity and bias perceptual report. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2081–2104. doi:10.1080/17470210902959279

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *Journal of Neurophysiology*, *101*(5), 2581–2600. doi:10.1152/jn.91198.2008

Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6996–7001. doi:10.1073/pnas.0502605102

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. doi:10.1016/j.neuroimage.2011.03.067

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, *24*(17), 2066–2070. doi:10.1016/j.cub.2014.07.060

Polosecki, P., Moeller, S., Schweers, N., Romanski, L. M., Tsao, D. Y., & Freiwald, W. A. (2013). Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *Journal of Neuroscience*, *33*(29), 11768–11773. doi:10.1523/JNEUROSCI.5402-11.2013

Popivanov, I. D., Jastorff, J., Vanduffel, W., & Vogels, R. (2012). Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *NeuroImage*, *63*(2), 723–741. doi:10.1016/j.neuroimage.2012.07.013

Popivanov, I. D., Jastorff, J., Vanduffel, W., & Vogels, R. (2014). Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *Journal of Neuroscience*, *34*(1), 95–111. doi:10.1523/JNEUROSCI.2748-13.2014

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, *18*(6), 2188–2199.

Qiu, D., Zaharchuk, G., Christen, T., Ni, W. W., & Moseley, M. E. (2012). Contrast-enhanced functional blood volume imaging (CE-fBVI): enhanced sensitivity for brain activation in humans using the ultrasmall superparamagnetic iron oxide agent ferumoxytol. *NeuroImage*, *62*(3), 1726–1731. doi:10.1016/j.neuroimage.2012.05.010

Rajimehr, R., Young, J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 1995–2000. doi:10.1073/pnas.0807304106

Reinl, M., & Bartels, A. (2014). Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics. *NeuroImage*, *102*, 407–415. doi:10.1016/j.neuroimage.2014.08.011

Rice, A., Phillips, P. J., & O'Toole, A. (2013). The role of the face and body in unfamiliar person identification. *Applied Cognitive Psychology*, *27*(6), 761–768. doi:10.1002/acp.2969

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, *24*(11), 2235–2243. doi:10.1177/0956797613492986

Richoz, A.-R., Jack, R. E., Garrod, O. G. B., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *65*, 50–64. doi:10.1016/j.cortex.2014.11.015

Roy, A., Shepherd, S. V., & Platt, M. L. (2012). Reversible inactivation of pSTS suppresses social gaze following in the macaque (Macaca mulatta). *Social Cognitive and Affective Neuroscience*, *9*(2), 209–217. doi:10.1093/scan/nss123

Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage*, *109*, 84–94. doi:10.1016/j.neuroimage.2015.01.012

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience*, *30*(39), 12978–12995. doi:10.1523/JNEUROSCI.0179-10.2010

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431. doi:10.1038/nn1786

Sacks, O. (1998). The man who mistook his wife for a hat. In *The Man Who Mistook His Wife for a Hat: and Other Clinical Tales* (pp. 8–22). New York: Simon and Schuster.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, *77*(1), 23–30.

Schmalzl, L., Zopf, R., & Williams, M. A. (2012). From head to toe: evidence for selective brain activation reflecting visual perception of whole individuals. *Frontiers in Human Neuroscience*, *6*, 108. doi:10.3389/fnhum.2012.00108

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, *194*(3), 465–475. doi:10.1007/s00221-009-1721-9

Schultz, J., Brockhaus, M., Bülthoff, H. H., & Pilz, K. S. (2013). What the human brain likes about facial motion. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(5), 1167–1178. doi:10.1093/cercor/bhs106

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience*, *25*(47), 11055–11059. doi:10.1523/JNEUROSCI.2621-05.2005

Singer, J. M., & Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *Journal of Neuroscience*, *30*(8), 3133–3145. doi:10.1523/JNEUROSCI.3211-09.2010

Sinha, P. (2011). Analyzing dynamic faces: key computational challenges. In C. Curio, H. H. Bülthoff, & M. A. Giese (Eds.), *Dynamic Faces* (pp. 177–185). MIT Press (MA).

Smirnakis, S. M., Schmid, M. C., Weber, B., Tolias, A. S., Augath, M., & Logothetis, N. K. (2007). Spatial specificity of BOLD versus cerebral blood volume fMRI for mapping cortical organization. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *27*(6), 1248–1261. doi:10.1038/sj.jcbfm.9600434

Song, Y., Luo, Y. L. L., Li, X., Xu, M., & Liu, J. (2013). Representation of contextually related multiple objects in the human ventral visual pathway. *Journal of Cognitive Neuroscience*, *25*(8), 1261–1269. doi:10.1162/jocn_a_00406

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience*, *17*(12), 1776–1783. doi:10.1038/nn.3855

Steede, L. L., Tree, J. J., & Hole, G. J. (2007). I can't recognize your face but I can recognize its movement. *Cognitive Neuropsychology*, *24*(4), 451–466. doi:10.1080/02643290701381879

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. doi:10.1038/nn.2163

Taubert, J., Van Belle, G., Vanduffel, W., Rossion, B., & Vogels, R. (2015). The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. *Journal of Neurophysiology*, *113*(5), 1644–1655. doi:10.1152/jn.00700.2014

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*(9), 989–995. doi:10.1038/nn1111

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670–674. doi:10.1126/science.1119983

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19514–19519. doi:10.1073/pnas.0809662105

Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. A. (2008). Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, *11*(8), 877–879. doi:10.1038/nn.2158

Van Essen, D. C. (2004). Organization of visual areas in macaque and human cerebral cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (pp. 507–521). Cambridge, MA: MIT Press.

Vanduffel, W., Orban, G. A., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., et al. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*, *32*(4), 565–577.

Vanduffel, W., Zhu, Q., & Orban, G. A. (2014). Monkey cortex through fMRI glasses. *Neuron*, *83*(3), 533–550. doi:10.1016/j.neuron.2014.07.015

Vangeneugden, J., De Mazière, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *Journal of Neuroscience*, *31*(2), 385–401. doi:10.1523/JNEUROSCI.2703-10.2011

Wachsmuth, E., Oram, M. W., & Perrett, D. I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex (New York, N.Y. : 1991)*, *4*(5), 509–522.

Watson, K. K., & Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Current Biology*, *22*(23), 2268–2273. doi:10.1016/j.cub.2012.10.016

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, *78*(1), 105–119.

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage*, *52*(4), 1559–1573. doi:10.1016/j.neuroimage.2010.04.262

Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *NeuroImage*, *56*(4), 2183–2199. doi:10.1016/j.neuroimage.2011.03.041

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97. doi:10.1007/s00426-011-0392-x

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. doi:10.3758/BRM.42.3.671

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, *5*, 633. doi:10.3389/fpsyg.2014.00633

Yovel, G., & Freiwald, W. A. (2013). Face recognition systems in monkey and human: are they the same thing? *F1000Prime Reports*, *5*, 10. doi:10.12703/P5-10

Yue, X., Pourladian, I. S., Tootell, R. B. H., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), E3467–E3475. doi:10.1073/pnas.1412616111

Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, *25*(36), 8150–8164. doi:10.1523/JNEUROSCI.2058-05.2005