Towards a unified perspective of object shape and motion processing in human dorsal cortex

Gennady Erlikhman, Gideon P. Caplovitz, Gennadiy Gurariy, Jared Medina, Jacqueline C. Snow

ARTICLE INFO

Keywords:
Dorsal stream
Action
Reaching
Grasping
3-D shape
Object perception
Shape perception
Spatiotemporal integration

ABSTRACT

Although object-related areas were discovered in human parietal cortex a decade ago, surprisingly little is known about the nature and purpose of these representations, and how they differ from those in the ventral processing stream. In this article, we review evidence for the unique contribution of object areas of dorsal cortex to three-dimensional (3-D) shape representation, the localization of objects in space, and in guiding reaching and grasping actions. We also highlight the role of dorsal cortex in form-motion interaction and spatiotemporal integration, possible functional relationships between 3-D shape and motion processing, and how these processes operate together in the service of supporting goal-directed actions with objects. Fundamental differences between the nature of object representations in the dorsal versus ventral processing streams are considered, with an emphasis on how and why dorsal cortex supports veridical (rather than invariant) representations of objects to guide goal-directed hand actions in dynamic visual environments.

1. Introduction

A core framework for our understanding of visual cognition is the division of cortical processing into two anatomically and functionally distinct pathways. In an initial conceptualization of the two visual pathways, which was based on lesion studies in monkeys, Mishkin & Ungerleider (1982) described a posterior-to-anterior hierarchical flow of visual information from early visual cortex that bifurcated into a ventral pathway extending anteriorly along lateral and ventral temporal cortex, and a dorsal pathway extending into posterior parietal cortex (PPC). According to this scheme, the ventral pathway is dedicated to processing object shape and identity (i.e., the ‘what’ pathway), and the dorsal stream is dedicated to processing the spatial position of objects (i.e., the ‘where’ pathway). Goodale and Milner (1992, 2018; Milner & Goodale, 1995) argued for a reformulation of this framework in which the emphasis was shifted from distinctions between visual inputs (i.e., cortical processing of visual object properties such as shape, versus object location), to distinctions between how visual information is used: for perception versus goal-directed action. In this framework, the dorsal stream must register accurate, in-the-moment, information about object shape, size and position and transform it into visuospatial coordinates appropriate for acting upon the object with the relevant effectors, such as the arm and hand. For the purpose of the current review, the important point is that in Goodale and Milner’s framework, object properties such as size and shape are computed in dorsal cortex in the service of controlling actions, rather than for guiding perception. For example, in their 1992
monograph entitled “Separate visual pathways for perception and action”, Goodale and Milner noted that damage to the dorsal stream “can impair the ability of patients to use information about the size, shape and orientation of an object to control the hand and fingers during a grasping movement, even though this same information can still be used to identify and describe the objects” (p. 21).

In line with this framework, dissociations between perception- versus action-related processes have been reinforced by over 30 years of convergent evidence from human psychophysics, neuroimaging, and lesion studies. In particular, the existence of object-selective perceptual representations in the ventral processing pathway has been documented extensively, from studies using neuroimaging (Grill-Spector et al., 1999; Kourtzi & Kanwisher, 2000, 2001; Kourtzi, Erb, Grodd, & Bülthoff, 2003; Malach et al., 1995), to studies of neuropsychological patients with visual agnosia, who, following lesions to ventral cortex, show severe and lasting deficits in their ability to recognize images of objects (Farah, Monheit & Wallace, 1991; Goodale, Milner, Jakobson, & Carey, 1991; Riddoch & Humphreys, 1987). Likewise, there has been widespread support for the notion that dorsal cortex supports vision for action. In addition to neuroimaging evidence (e.g., Bälint, 1909; Holmes, 1918; Jakobson, Archibald, Carey, & Goodale, 1991; Karnath & Perenin, 2005; Perenin & Vighetto, 1988; Pierrot-Deseille, Rivaud, Gaymard, & Agid, 1991; Ratcliff & Davies-Jones, 1972; Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997), the role of the dorsal pathway in mediating vision for action has been extensively corroborated by neuroimaging (Clower et al., 1996; Culham, Cavina-Pratesi, & Singhal, 2006; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Kawashima et al., 1996) and monkey neurophysiology (Colby, Duhamel, & Goldberg, 1993; Greffes & Fink, 2005; Johnson, Ferraina, Bianchi, & Caminiti, 1996; Murata, Gallese, Luppino, Kaseda, & Sakata, 2000; Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997) studies.

Since Goodale and Milner’s original work, object-selective representations were discovered in PPC of the dorsal processing pathway, first in non-human primates (Denys et al., 2004; Lehky & Sereno, 2007; Murata et al., 2000; Sereno & Maunsell, 1998; Taira, Mine, Georgopoulos, Murata, & Sakata, 1990), and later in humans (Denys et al., 2004; Fang & He, 2005; Koen & Kastner, 2008; Mruczek, von Loga, & Kastner, 2013). Interestingly, these dorsal visual object areas appear to be recruited independently of attention, and irrespective of whether a grasping movement was planned or initiated by the observer. In one of the first detailed studies of object representations in human dorsal cortex, Koen and Kastner (2008) used an fMRI-adaptation paradigm to measure the selectivity profile of different regions of human cortex to object shape, as well as to object size and viewpoint. Observers passively viewed black-and-white computerized images of objects, such as silhouettes of basic geometric shapes, line-drawings of everyday familiar objects and tools, and images of objects whose 3-D shape was defined by monocular depth cues such as shading, specular highlights and occlusion. Within the ventral stream, neural populations selective for shape included intermediate (e.g., V4) and higher-level areas such as the lateral occipital complex (LOC). Importantly, object-selective responses were also observed within topographically-defined areas along the dorsal pathway, including V3A, MT, and V7, as well as in ventral regions of PPC in the intraparietal sulcus (IPS), in IPS1 and IPS2. Moreover, both LOC and IPS1 and IPS2, responded similarly, despite changes in the size and orientation of the images, suggesting that object areas in inferotemporal cortex and PPC share a pattern of increasing invariance (or generalization) across changes in viewing conditions. Based on these results, Koen and Kastner (2008) argued that basic information about object shape, size and viewpoint, are represented in the ventral and dorsal pathways, and that the representations in each stream appear to be strikingly similar.

Since these early reports of shape processing in the dorsal stream, research into the nature of object-related processing in dorsal cortex has been burgeoning. There has been a rapid increase in studies documenting the cortical organization of object-related representations using fMRI (Chandrasekaran, Canon, Dahmen, Kourtzi, & Welchman, 2007; Erlikhman, Gurary, Mruczek, & Caplovitz, 2016; Ludwig, Kathmann, Sterzer, & Hesselmann, 2015) and the temporal aspects of dorsal object processing using EEG (Wokke, Scholte, & Lamme, 2014; Zaretskaya & Bartels, 2015; Zaretskaya, Anstis, & Bartels, 2013). Others have emphasized the involvement of dorsal cortex (in addition to ventral stream areas) in processing particular classes of objects, such as man-made tools (e.g., Chao and Martin, 2000; Lewis, 2006; Macdonald and Culham, 2015). However, the extent to which these dorsal object representations reflect ‘toolness’ versus other attributes of the stimuli, such as their elongation, is still a subject of debate (Almeida et al., 2014; Chen, Snow, Goodale, & Culham, 2017; Sakuraba, Sakai, Yamanaka, Yokosawa, & Hirayama, 2012). These recent neuroimaging studies have been complimented by case studies in neuropsychological patients, for whom aspects of object processing remain intact despite the fact that contributions to object processing from the ventral stream are reduced or eliminated due to brain damage (Frend, Rosenthal, Ganel, & Avidan, 2015). The data from these neuropsychological patients suggest that dorsal cortex may process shape during perception, in addition to action-related tasks (Milner et al., 1991; Goodale et al., 1991). Nevertheless, in line with Goodale and Milner’s (1992) two visual systems model, this shape processing was insufficient to support conscious awareness because the agnosia patients in Freud et al.’s (2015) study were unable to perform explicit shape-related judgments above-chance. A recent study using the continuous flash suppression approach, which effectively suppresses object processing in the ventral- but not the dorsal pathway, further supports these earlier patient findings by demonstrating that 3-D shape representations computed in the dorsal pathway can contribute to perceptual decisions (Fang & He, 2005; Freud, Robinson, & Behrmann, 2018).

Although much has been learned about dorsal cortex over the past 20 years, there remain a number of outstanding questions and controversies. The functional significance of dorsal stream object representations is unclear, and a detailed account of object processing in dorsal cortex is lacking (Frend, Plaut, & Behrmann, 2016; Kravitz, Saleem, Baker, & Mishkin, 2011). Important unresolved questions include whether object information is computed directly within neural populations in parietal cortex (or is relayed to dorsal cortex, perhaps via areas V3 and/or V4), what types of object properties are processed along the dorsal pathway, and the extent to which the dorsal and ventral pathways represent redundant or unique object information (Koen & Kastner, 2008). Despite Milner and Goodale’s (1992) emphasis on the fact that the output requirements for visually-guided actions with objects requires constant in-the-moment updating of object properties and spatial position in egocentric space, research in the domain of 3-D shape processing, and spatiotemporal integration, in dorsal cortex has unfolded largely in parallel. For example, with respect to what types of object...
properties are processed in the dorsal pathway, some have argued that dorsal cortex is critically involved in processing 3-D geometric shape information, such as surface curvature, size, and location relative to the observer (Farivar, 2009; Freud et al., 2016; Janssen, Verhoef, & Premereur, 2018; Orban, 2011). A separate branch of research has focused on the involvement of dorsal cortex in motion processing and form-motion interactions (Galletti & Fattori, 2018; McCarthy, Erlikhman, & Caplovitz, 2017). In the sections below, we relate research outcomes in 3-D shape representation and form-motion integration in dorsal cortex, with respect to their contribution to action-related object processing. We also revisit the ways in which dorsal object processes differ in fundamental ways from the known properties of object areas along the ventral processing stream (DeYoe & Essen, 1988; Goodale & Milner, 1992). A synthesis of recent research in each of these domains will foster a more complete understanding of the nature of object processing in the dorsal visual stream, and serve as a guide to future research in the field.

2. Hallmarks of object-related representations in dorsal cortex

Naturalistic behavior in real-world environments is a computationally demanding task that involves not only perceiving and recognizing objects, but also grasping and interacting with those objects. Consider the example of a coffee cup sitting on a cluttered desk (Fig. 1). To recognize the cup and tell it apart from other objects sitting on the desk, its shape must be computed, together perhaps with information about its relative size (Konkle & Oliva, 2012), and material properties (Cant & Goodale, 2006). In fact, discriminating a cup from other objects (such as a pen or a key) is a perceptual task that can be achieved with surprisingly impoverished visual inputs, such as when only a silhouette outline of the object is available due to low illumination. However, in order to determine whether, when, and how, to pick it up and what to do with it afterward, additional visual metrics are required – ones that are fundamentally different from those required for recognition (Goodale & Milner, 1992; Milner & Goodale, 1995, 2008; Thaler & Goodale, 2010). For example, a distinct 3-D volumetric representation of the shape, size and orientation of the cup is necessary to plan how best to open one’s hand and shape the fingers for a grasp. There also needs to be information about the cup’s location in egocentric space (i.e., its position in 3-D space relative to the observer) to know where to reach, as well as its position in depth relative to other objects on the table to avoid knocking them over when making a movement toward the cup. Importantly, the objects

Fig. 1. The Dorsal Stream in Action. (A). Consider the challenges of performing a simple visuomotor task, such as pouring orange juice from a bottle on the table, into a glass. The observer must determine the location, distance, and size of the object relative to himself. Initiating a reach towards the object involves constant monitoring and updating of hand position to avoid knocking over other objects (i.e., obstacles) that may be in the way. He must also determine the 3-D geometric structure of the object, so that the fingers can be shaped appropriately to match the anticipated contact site. To do pour contents from (rather than simply move) the bottle effectively, he needs to apply a grip force that is commensurate with the object’s shape and weight. This weight estimate presumably requires information about the object’s size and material properties – information that needs to be recalibrated on-the-fly, given that the object’s weight will change as liquid moves from the bottle to the glass. (B). These computations become even more complex when the objects we wish to interact with are in motion. For example, we may notice a cup falling from the table, and instinctively reach to grasp it. This action requires many of the dynamic computations described above (i.e., reaching in the right direction and adjusting the fingers to match the object’s shape), but now they must be applied to an object whose spatial position and distance from the observer are changing rapidly. Both 3-D structure, and form-motion integration, are required to interact with objects in dynamic real-world contexts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
we interact with in naturalistic environments often change position and move with variable velocity over time. As dynamic objects (and observers) move through the environment, their relative positions also differ with respect to retinotopic (eye-centered), egocentric (body-centered) and world-based coordinate reference frames (Colby & Goldberg, 1999; Galletti & Fattori, 2018). Below, we review converging evidence from studies in monkeys and humans that has identified a network of areas in dorsal visual cortex that are sensitive to many of these 3-D shape-, position- and velocity-related properties that are necessary to successfully interact with objects in our environment.

2.1. Information about static objects

2.1.1. 3-D shape selectivity

There are many visual cues, including stereopsis, texture, shading, and motion, that the visual system uses to recover depth information from the input on the retina, in the service of reconstructing 3-D shape information (for more detailed reviews see Todd, 2004, and Orban, 2011). Cortical areas sensitive to 3-D depth cues form a broad neural network that extends across the cerebral mantle, including early visual areas (V1-V3) where individual features such as luminance, stereopsis, and texture are locally detected, intermediate ventral visual areas such as V4 where they likely begin to be integrated, as well as areas farther along the ventral processing stream, such as the lateral occipital cortex (LOC) and ventral temporal cortex (Moore & Engel, 2001; Verhoeft, Vogels, & Janssen, 2010, 2012; Janssen, Vogels, & Orban, 1999, 2000; Grill-Spector & Weiner, 2014). Dorsal cortex in particular appears to be intimately involved in processing higher-level second-order (i.e., slant) and third-order (i.e., curvature) depth information about 3-D object structure. Areas that encode higher-order depth features of objects in dorsal cortex include hMT+, MST, V3A/B, V6, V7, and regions along the intraparietal sulcus (IPS) (Alizadeh, Van Dromme, Verhoeft, & Janssen, 2018; Georgieva, Peeters, Kolster, Todd, & Orban, 2009; Janssen et al., 2000; Katsuyama, Usui, Nose, & Taira, 2011; for reviews see Anzai & DeAngelis, 2010; Freud et al., 2016; Janssen et al., 2018; Orban, 2011; Theys, Romero, van Loon, & Janssen, 2015; Tsao, Conway, & Livingstone, 2003; Tsao, Vanduffel, et al., 2003; Tsutsui, Taira, & Sakata, 2005). For example, in monkeys the caudal part of the intraparietal area (CIP), which corresponds to VIPS, V7/IPSO, or V7A in humans (Konen, Mruczek, Montoya, & Kastner, 2013; Orban, 2016), is involved in representing 3-D curvature (Alizadeh et al., 2018; Georgieva et al., 2009; Janssen et al., 2000; Katsuyama et al., 2011). Stereoscopically defined pairs of 3-D shapes that differ only in their sign of curvature (i.e., concave versus convex) selectively activate individual neurons in this area (Janssen et al., 2000). In an fMRI study in humans using the same stimuli, Georgieva et al. (2009) found that such surfaces preferentially activate dorsal areas V3A, V7, IPS as well as the posterior inferior temporal gyrus, compared to flat and monocularly viewed surfaces without 3-D structure. Furthermore, temporary inactivation of area CIP in monkeys leads to impairments in perceiving depth structure and 3-D curvature, attesting to its critical role in representing those 3-D object properties (Van Dromme, Premereur, Verhoeft, Vanduffel, & Janssen, 2016). Other studies suggest that CIP may also be a locus where different types of depth cues are integrated, including shape-from-shading and texture (Ban, Preston, Meeson, & Welchman, 2012; Murphy, Ban, & Welchman, 2013; Nelissen et al., 2009; Tsutsui, Jiang, Yara, Sakata, & Taira, 2001). CIP may be part of a larger object-processing network. In monkeys, CIP projects to anterior regions of the intraparietal area (AIP) – an area involved in object grasping, and which in turn projects to premotor area F5 (Theys, Pani, van Loon, Goffin, & Janssen, 2012, 2013) (for reviews, see Janssen et al., 2018; Theys et al., 2015). AIP, like CIP, is also sensitive to 3-D curvature and 3-D shape, (Durand, Celebrini, & Trotter, 2007; Durand, Nelissen, et al., 2007; Durand, Peeters, Norman, Todd, & Orban, 2009; Joly, Vanduffel, & Orban, 2009; Murata et al., 2000; Srivastava, Orban, De Mazière, & Kanssen, 2009; Verhoeft et al., 2010). CIP also projects to inferotemporal cortex (IT), and it has recently been suggested that 3-D shape signals in ventral areas may be driven ultimately by parietal processing (Van Dromme et al., 2016). In humans, detailed studies of isolated neural populations are lacking. However, neuropsychological evidence indicates that patients with ventral cortex damage show severe impairments in tests of explicit object recognition, they can nevertheless judge whether one part of a 3-D shape is nearer or farther away than another part of the shape (Freud et al., 2015).

2.1.2. Sensitivity to object size and distance

In addition to recovering detailed information about the 3-D geometric depth structure of objects, object-related areas in dorsal cortex appear to be unique (compared to ventral cortex) with respect to their coding of object size. The physical size of an object is likely to be important, not only for planning and executing actions with objects, but also for assessing action affordances even when there is no explicit plan to grasp (Gomez, Skiba, & Snow, 2018; Tucker & Ellis, 2001). Convergent evidence for the notion that dorsal cortex is involved in the size coding of graspable objects comes from human neuroimaging and monkey neurophysiology. Murata et al. (2000) recorded from cells in the anterior intraparietal area (AIP) while monkeys either passively viewed or grasped solid 3-D objects. The objects were small, medium, or large in size. More than half of the ‘visually-responsive’ neurons responded when the monkey fixated the object (‘object-type’ neurons), and most of these cells were tuned for both the shape and size of the object. Importantly, however, the visual-motor neurons in AIP were intermixed with motor-dominant and visuo-motor neurons, suggesting a role for this area in visuo-motor transformations (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995; Sakata, Tsutsui, & Taira, 2005). In humans, topographic representations of object size have been identified in bilateral human parietal cortex using ultra-high-field (7T) functional MRI (Harvey, Fracasso, Petridou, & Dumoulin, 2015). Levels of GABA (the primary inhibitory neurotransmitter in the human brain) in parietal cortex have also been shown to selectively influence size perception (Song, Sandberg, Andersen, Blicher, & Rees, 2017).

To accurately represent the size of an object for the purpose of grasping, it is also necessary to have accurate metrics of the object’s distance, since objects of different sizes can project the same 2-D image on the retina, depending on the distance at which they are viewed (Murray, Boyaci, & Kersten, 2006). Dorsal cortex plays a unique role in representing an object’s position in depth. One of the
most powerful cues to position in depth is stereoscopic disparity (Howard & Rogers, 1995). In addition to early visual areas, several regions of monkey parietal cortex, including LIP and AIP, are sensitive to disparity information and subtle changes in the apparent depth of a stimulus (Gonzalez & Perez, 1998; Janssen, Srivasta, Ombelet, & Orban, 2008; Theyes et al., 2015). In humans, depth processing is localized to a number of areas including V3A (Backus, Fleet, Parker, & Heeger, 2001; Berryhill & Olson, 2009; Tsao, Conway, et al., 2003; Tsao, Vanduffel, et al., 2003), V7/IP50 (Brouwer, van Ee, & Schwarzbach, 2005), and other areas along the caudal intraparietal sulcus (Rutschmann & Greenlee, 2004). It should perhaps not be surprising that the same areas that are sensitive to 3-D shape information in both monkeys and humans as discussed in the previous section are also implicated in the processing of relative 3-D information.

Although such relative distance information can serve as a cue to object size, knowing the absolute distance of an object from the body (i.e., egocentric distance) is critical for an active observer to determine whether the object lies within reach of the effector (Jeannerod, 1981) and to determine the appropriate hand configuration for grasping (Castiello, 2005; Jeannerod, 1986; Smeets & Brenner, 1999). In both primates and humans, parietal cortex is tiled with areas that are specialized for actions within the co-ordinate frames of different effectors, including the eyes, arms, and hands (Andersen & Buneo, 2002; Andersen, Snyder, Bradley, & Xing, 1997; Colby & Goldberg, 1999; Gallivan & Culham, 2015). Some of the dorsal areas tuned to stereo-depth in humans show greater sensitivity to egocentric rather than allocentric distance, that is, to distance between the observer and the object rather than the distance between objects or surfaces in the environment (Neggers, Van der Lubbe, Ramsey, & Postma, 2006; Nerri, Bridge, & Heeger, 2004). The region of space around the body that lies within reach (often known as peripersonal space) is prioritized in many areas of dorsal (Duhamey, Bremmer, Ben Hamed, & Graf, 1997; Galletti, Battaglini, & Fattori, 1993) and frontal (Fogassi et al., 1996) cortex. Intuitively, this makes sense because our arms and hands can only operate within a certain anatomical range.

In humans, an area within the dorsal pathway, known as the superior parieto-occipital sulcus (SPOC), responds strongly to solid reachable 3-D objects, even when participants are not required to perform a reaching movement to the object (Cavina-Pratesi et al., 2018). SPOC in humans most likely corresponds with areas V6 and V6A in the monkey which, in turn, project to dorsal premotor cortex (Galletti & Fattori, 2018; Galletti, Kutz, Gamberini, Breveglieri, & Fattori, 2003). Responses to solid objects in SPOC are also modulated by the egocentric distance of the object. Using fMRI, Gallivan, Cavina-Pratesi, and Culham (2009) presented right-handed human observers with small Lego shapes on a platform. The stimuli were positioned on the platform so that they fell either within or outside of reach of the observer. On some trials, participants were asked to view the objects without touching them, while on other trials their task was to reach to touch or grasp the objects. fMRI activation in SPOC in the left hemisphere (contralateral to the dominant hand) was significantly higher for passively-viewed objects that were positioned within versus outside of reach. Given that SPOC responded to reachable objects, even when no action was required, the authors argued that activation in SPOC reflects automatic processing of motor affordances (Gibson, Owlsley, Walker, & Megaw-Nyce, 1979). Interestingly, SPOC appears to represent more than just egocentric distance, but also previous visuomotor experience and manual predispositions. Such predispositions are commonplace in everyday life, as evidence by the motor behaviors of left- versus right-handers. Left-handers typically use both hands to manipulate objects, whereas right-handers tend to rely on the dominant, right hand alone (Gonzalez & Goodale, 2009). Constraints from action, such as hand preference, can influence attentional selection of objects, even when there is no explicit plan to grasp the stimulus (Gomez et al., 2018; Gomez & Snow, 2017; Humphreys et al., 2010). Accordingly, Gallivan, McLean, and Culham (2011) compared fMRI responses to graspable objects in left- and right-handed observers, and found that for right-handed observers, SPOC responded most strongly for objects that were within reach of the right hand. Conversely, for left-handers, SPOC responded to objects within reach of both the right and left hands. Convergent evidence for a role of dorsal cortex in encoding objects within egocentric space comes from EEG studies of desynchronization of the $\mu$ (‘mu’) rhythm. Mu rhythms are neural signatures of sensorimotor processing recorded over centro-parietal electrodes; desynchronization of the $\mu$ rhythm is understood to reflect neural processes involved in automatic action preparation (Pineda, 2005). Using realistic stimuli presented in the context of virtual reality, Wamain, Gabrielli, and Coello (2016) reported significantly stronger $\mu$ desynchronization for objects that were positioned within versus outside of reach. Critically, the effect of distance on $\mu$ desynchronization was only apparent when the stimulus formed a coherent shape, but not when the image depicted a scrambled version of the object.

The finding that occipito-parietal areas may be the locus of egocentric distance, size, and 3-D shape information integration is supported by a report of a patient with bilateral occipito-parietal damage (Berryhill, Pendrich, & Olson, 2009). This patient was unable to use either stereo or monocular cues such as occlusion, shading and linear perspective to recover the shapes of objects. For example, in displays with two overlapping simple shapes, like a square and a circle, she had difficulty in saying which one was closer. When shown pictures of volumes and 2-D shapes, she described cubes as squares, cylinders as circles, and pyramids as triangles. This patient also had difficulty in both estimating the distance to nearby objects and in placing those objects at specific distances, especially if the objects were unfamiliar; however, she showed relatively little impairment when asked to estimate the distance between objects or to place one object at a certain distance from another. That is, her egocentric distance perception was impaired, but her allocentric or relative distance perception remained intact. In addition to impairments in perceiving shape and distance, she also exhibited impairments in size perception. For example, she was slightly impaired in comparing the size of objects that were matched for retinal size, and greatly impaired in comparing objects that were matched in physical, but not retinal size, often reporting that the object with larger retinal size were physically larger. When either retinal or physical size was held constant, she had difficulty in reporting relative distance between the objects. The impairment was stronger when she viewed the stimuli monocularly. For this patient, relatively restricted, bilateral occipito-parietal lesions simultaneously impaired 3-D shape perception, egocentric distance perception, and size perception.

Taken together, these behavioral and neurophysiological findings point to a network of parietal and dorsal areas including V3A, V7/IP50, IPS and SPOC (and the corresponding monkey homologs), that are involved in processing object properties that are
important for manual interaction. The functions of these areas are likely to include the representation of 3-D shape, object size, relative 3-D distance between the surfaces of an object, position in depth, and distance between the observer and the object, perhaps with special importance placed on objects that are within reach. One important feature of dorsal representations, which we will return to in later sections, is their non-invariant nature. Unlike in the ventral visual processing stream, which can be characterized broadly as leading towards abstraction, categorization, and recognition of objects, these dorsal areas appear to represent veridical, metric information that is critical for planning and executing actions with objects.

### 2.2. Information about moving objects

When studying shape, depth, and distance perception in the lab using unobstructed, static, 2-D images, it is easy to overlook the fact that in natural environments we must often interact with 3-D, dynamic stimuli that are only partially visible at any given moment. In order to accurately recover the 3-D shape of dynamic objects in environments through which the observer themselves may be moving, the visual system must distinguish self- from object-motion and integrate motion and form information over both space and time. Consider the case of a parent trying to pick their child who is running through a crowd. Information about the child’s 3-D shape, size, absolute position in depth, and their egocentric position, although critical, are likely to arrive piecemeal and be partial and incomplete given the busy scene and dynamic visual input. For example, as they move among the crowd only parts of the child may be visible at any moment, yielding a partial representation of overall shape, which itself is changing in position moment-to-moment. A consideration of the nature of shape processing in more complex, dynamic visual environments, together with the known properties of object-selective areas in ventral cortex (which are devoted primarily to invariant representations), suggests a role for dorsal cortex in representing both visual motion and shape as well as visuo-motor representations.

A critical, early step in processing object motion is to distinguish motion signals that arise from an object moving through space from those that are the result of self-motion. In the example of the parent chasing a child, the motion signals available at the retinae are a combination of signals that arise from the parent turning their head, moving their eyes, moving their body through the environment, and the motion of the child. Intriguingly, many regions of dorsal cortex that are implicated the coding of 3-D shape, such as V3A, V6, V6A and regions along IPS, are also known to be involved in global motion perception (for reviews, see Galletti & Fattori, 2018; Gilaine-Dotan, 2016; Kravitz et al., 2011; Zeki, 2015). In fact, there is a tight relationship between motion and shape processing (for a review, see Mather, Pavan, Marotti, Campana, & Casco, 2013). Consideration of such form-motion interactions is important for understanding the terminal goals of visual processing, such as recognition and action, because the extraction of both motion and shape information is a necessary intermediate step toward constructing veridical shape representations from dynamic information. Below, we consider two broad categories of form-motion interactions: influences of form on motion perception and influences of motion on form perception.

#### 2.2.1. The effects of form-motion interactions on perception

Form and motion processes can mutually interact with each other to influence both if and how an object appears to move as well
as the perceived shape of the object as it moves. Each of these can be revealed by a wide range of perceptual phenomena that reveal just how diverse these interactions are. One example of form influencing motion is Transformational apparent motion (TAM). In TAM observers see a sequence of static images that induce the experience of apparent motion (Hikosaka, Miyachi, & Shimojo, 1993a, 1993b; Hsieh, Caplovitz, & Tse, 2006; Kanizsa, 1951; Tse & Caplovitz, 2006; Tse & Logothetis, 2002; Tse, Cavanagh, & Nakayama, 1998). Unlike most apparent motion displays, however, in TAM, the perceived motion is seen in a direction that preserves the continuity of their contours instead of to the spatiotemporally nearest positions. That is, how an object is perceived to move is determined by the relationship of its contours at one position and time to the shape presented at another time, and not merely by where that shape appears.

Another example of how form can influence motion perception, the reverse is also true. For example, eye movements may make tilted lines appear vertical (Bridgeman, Mayer, & Glenn, 1976). A moving background may change the apparent size, shape and position of briefly flashed objects (Anstis & Cavanagh, 2017). Movement speed can affect perceptual grouping and global shape (Anstis, 2005), and, in a now well-studied occluded-aperture stimulus, stimulus configuration can affect whether four bars appear to be a single object moving coherently in one direction or as four separate objects moving independently (Lorenzoceu & Alais, 2001; Lorenzoceu & Shiffrar, 1992; McDermott & Adelson, 2004; Shiffrar & Lorenzoceu, 1996; Tang, Dickinson, Visser, Edwards, & Badcock, 2015). Rotational motion in particular has been found to distort the perceived shape and size of objects (Anstis, Sturzel, & Spillmann, 1999; Caplovitz & Tse, 2006; McDermott & Adelson, 2004; Meyer & Dougherty, 1990; Shiffrar & Pavel, 1991; Stanley, 1968; Weiss & Adelson, 2000). When a square rotates, for example, they can appear as pin cushions with concave sides potentially caused by perceived compression along the curved motion path (Anstis & Kaneko, 2016). However, not all form-motion interactions result in illusions or misperceptions of one or the other feature. In some cases, one of the features can be used to produce the other. For example, in the kinetic depth effect, the 2-D projection of the rotation of a wire-frame object produces a vivid perception of 3-D structure (Wallach & O’Connell, 1953). Similarly, in structure-from-motion, unconnected elements moving along a 3-D surface can lead to the perception of that surface (Ullman, 1979). Just as shading, texture, and stereopsis can provide 3-D shape information, motion signals from spatially sparse elements can be integrated over time to produce a 3-D shape representation.

There is converging evidence that the neural correlates of form-motion interactions include several dorsal areas including V3A, V3B/KO, MT/V5, MST, and also IPS in humans (Beer, Watanabe, Ni, Sasaki, & Andersen, 2009; Murray, Olshausen, & Woods, 2003; Vanduffel et al., 2002). For example, transformational apparent motion (TAM) and rotational motion of trackable features have both been shown to activate area V3A (Caplovitz & Tse, 2007b; Tse, 2006). As these stimuli involve the perception of motion, it is perhaps not surprising that these cortical areas are also sensitive to motion and optic flow patterns (Kononen et al., 2003; Sugihara, Murakami, Sheny, Andersen, & Komats, 2002). However, when the motion signal is not just a field of moving dots, but can be used to define a form, as in structure-from-motion, V3A is more strongly activated and this activation gradually increases along the dorsal pathway along IPS (Caclin et al., 2012; Klaver et al., 2008; Orban, Sunaert, Todd, Van Hecke, & Marchal, 1999; Paradis et al., 2000; Peuskens et al., 2004; Vanduffel et al., 2002). In humans as compared to monkeys, areas V3A and IPS seem to be more involved in perceiving motion-defined shape, whereas other areas such as MT/V5 respond to both structured motion stimuli and global motion patterns that lack form (Orban et al., 2003; Paradis et al., 2000). Similarly, cortical lesions in IPS lead to impairments in perceiving global motion patterns (Lestou, Lam, Humphreys, Kourtzi, & Humphreys, 2014), while TMS applied to IPS biases perception toward local motion over global grouping and motion (Zaretskyai et al., 2013) and disrupts global motion processing (Cai, Chen, Zhou, Thompson, & Fang, 2014). Perhaps, in addition to integrating depth and spatial information to produce observer-relative 3-D shape representations, these intermediate and dorsal areas such as IPS also perform temporal integration of shape information.

### 2.2.2. Spatiotemporal form integration

In the preceding section, we considered how the accurate recovery of shape, velocity, and position information may be affected by other object or scene properties. However, in most of our examples, we considered stimuli that were completely visible at any given time. In contrast, in natural environments, it is not uncommon to have our view of an object obstructed by an occluder. As we move through the world (or as other objects move), the input to our visual system is fragmented, both in space and time, so that at any given moment we may only see part of a shape. Therefore, in addition to processing 3-D object shape and motion, the visual system must also integrate spatiotemporally segregated information about object parts into coherent perceptual units (Kellman, Garrigan, Shipley, Yin, & Machado, 2005; McCarthy et al., 2017; McCarthy, Strother, & Caplovitz, 2015; Palmer, Kellman, & Shipley, 2006). A number of behavioral studies have demonstrated that the human visual system is capable of constructing whole object percepts from dynamic and partial visual information. For example, illusory edges and figures can be seen even when the luminance-defined edges
that support them are only shown one at a time, or are gradually revealed through patterns of occlusion and disocclusion (Andersen & Cortese, 1989; Erlikhman & Kellman, 2016a, 2016b; Gibson et al., 1979; Kaplan, 1969; Kellman & Cohen, 1984; Kellman & Loukides, 1987; Kojo, Liinasuo, & Rovamo, 1993; McCarthy et al., 2015; Palmer et al., 2006; Shipley & Kellman, 1993, 1994, 1997; Stappers, 1989; Yonas, Craton, & Thompson, 1987).

McCarthy et al. (2015a) created a display that nicely demonstrates form-motion integration. The authors used a Kanizsa display in which four circles with sectors removed were arranged around the corners of a square. Classically, when this stimulus is shown in static form, the percept is of an illusory white square on top of four black circles (Kanizsa, 1979). In McCarthy et al.’s (2015) study, the illusory square rotates by changing the angle of the cutout in the inducing elements from frame to frame. Critically, each of the four inducers did not appear simultaneously, but, rather, they appeared one at a time. McCarthy et al. found that, under these conditions, the perception is of a rigidly rotating square. This implies that, in between inducer presentations, the represented position of the square was updated to match the shape of the cutout of the next inducer to appear. In short, in order to accurately construct a percept of a rigidly rotating object from partial information, the visual system must (1) maintain position and velocity information of no-longer-visible object parts, (2) update their positions when they are no longer visible, and (3) integrate previously visible with currently visible surfaces (McCarthy et al., 2017; Palmer et al., 2006). These processes occur regularly in the recovery of shape information in dynamic environments and are critical if one aims to interact with moving objects in the world.

Several studies have isolated neural correlates of spatiotemporally dynamic stimuli in areas along the dorsal stream, including V3A, V3B/KO, hMT+, and the IPS, many of which are implicated in 3-D shape processing, spatial localization, and global motion processing (Erlikhman et al., 2016; Erlikhman & Caplovitz, 2017; McCarthy, Kohler, Tse, & Caplovitz, 2015; Reichert et al., 2014; Zaretskaya et al., 2013; Zaretskaya & Bartels, 2015). For example, dynamic Kanizsa figures, such as the ones described above, activate area V3A only when those fragments are perceived as unified into a single object (McCarthy et al., 2015). Dynamically-defined form information (e.g. shape or overall orientation) can be decoded in both V3A and IPS from contours that are only revealed gradually over time, as when seen through an aperture (Erlikhman et al., 2016; Kuai, Li, Yu, & Kourtzi, 2016). These areas may be part of the same integrating network (Perry & Fallah, 2014): the perception of global structure in some classes of stimuli elicits rapid feedback from IPS to V3A 100–200 ms after stimulus presentation (Liu, Wang, Zhou, Ding, & Luo, 2017). Stimulation applied to V3B disrupts feedback to earlier visual areas and prevents such contour integration (Li, Wang, & Li, 2017). These same areas – V3A/B and IPS – have also been implicated in the maintenance of no-longer-visible object information (Erlikhman & Caplovitz, 2017; Fang & He, 2005; Weigelt, Kourtzi, Kohler, Singer, & Muckli, 2007), a key feature necessary for the construction of spatiotemporal representations.

3. Summary and conclusions

The accumulating empirical evidence indicates that object properties such as form and motion are represented in dorsal cortex. Importantly, although object information is represented across both visual streams, the representations in dorsal cortex are unlikely to be redundant with those in ventral cortex (Konen & Kastner, 2008). Rather, dorsal object-representations are unique from those in ventral cortex in several fundamental respects. First, object processing in dorsal cortex follows an overarching posterior-to-anterior visual-to-motor gradient, with more posterior areas (such as AIP) dedicated predominantly to analysis of visual properties such as 3-D curvature, while anterior areas (such as AIP) are increasingly concerned with computations related to potential motor interactions with the stimulus, such as encoding the object’s absolute object location in egocentric space, reaching trajectory, and hand pre-shaping (Culham and Valyear, 2006; Fabbri, Stubbs, Cusack, & Culham, 2016; Shmuelof and Zohary, 2005; Stark & Zohary, 2008). In fact, this visual-to-motor gradient may not be limited to the rostro-caudal axis, but may also extend along a medio-lateral axis as well, reflecting the underlying connectivity patterns with the visual and motor cortices (Freud et al., 2016; Kravitz et al., 2011). Second, object-related areas of dorsal cortex appear to be devoted to computing veridical (rather than invariant) representations of objects. Whereas shape responses in areas of ventral cortex show a global pattern of increasing tolerance to changes in visual attributes such as object location and size (Ito, Tamura, Fujita, & Tanaka, 1995; Op de Beeck & Vogels, 2000; Schwartz, Desimone, Albright, & Gross, 1983; Vuilleumier, Henson, Driver, & Dolan, 2002), occlusion (Kovacs, Vogels, & Orban, 1995), and viewpoint (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011; Grill-Spector et al., 1999; Grill-Spector, Kourtzi, & Kanwisher, 2001; James, Humphrey, Gati, Menon, & Goodale, 2002; Vuilleumier et al., 2002), this does not appear to be the case in dorsal cortex. Rather, as we have described above, neural populations across dorsal cortex are essentially sensitive to changes in features such as 3-D shape, orientation, slant, and absolute distance relative to the body and effectors. These populations are also sensitive to changes in an object’s position over time (Janssen et al., 2008; Theys et al., 2015). In other words, whereas ventral cortex is devoted to forming increasingly abstract representations of shape that remains stable across changes in the visual information on the retina, dorsal cortex supports a different representation of objects that is sensitive to minor stimulus transformations.

Although we have emphasized the unique role of dorsal cortex in representing 3-D shape, location, distance, and object motion, shape-selective regions within the ventral and dorsal pathways nevertheless seem to interact during shape perception (Galletti & Fattori, 2018; Janssen et al., 2018; Van Dromme et al., 2016; Wokke et al., 2014). A number of regions in dorsal cortex (such as AIP and CIP) share anatomical connections with object-selective areas in the ventral pathway (Borra et al., 2008; Liu et al., 2017; Premereur, Van Dromme, Romero, Vanduffel, & Janssen, 2015; van Polanen & Davare, 2015; Webster, Bachevalier, & Ungerleider, 1994; Yeatman et al., 2014). In humans, these large-scale networks may share a central hub in intermediate area V3A (Liu et al., 2017). Importantly, however, the extent to which ventral or dorsal areas are recruited during shape perception may be heavily task-dependent. For example, when observers are asked to report a location change across multiple presentations of an object, both ventral (LOC) and dorsal (PPC) regions are also sensitive to simultaneous changes in the object’s shape and identity, even though location
was the only task-relevant property (Zachariou, Kltazky, & Behrmann, 2014). In contrast, when the task is to notice a shape difference across repeated presentations and location is irrelevant, only dorsal areas show location sensitivity. Similarly, face stimuli defined by structure-from-motion activate both ventral and dorsal regions, but the ventral regions are activated only when observers are asked to perform a recognition task (Farivar, Blanke, & Chaudhuri, 2009; Kriegeskorte et al., 2003). Greater activation of dorsal areas during a recognition task has also been shown to predict activity in ventral areas (Sim, Helbig, Graf, & Kiefer, 2015).

The converging evidence for veridical (rather than invariant) object representations in dorsal cortex seems to contrast with early fMRI findings, in which object responses in dorsal cortex were thought to reflect a pattern of invariance, similar to that of ventral object areas such as LOC (Konen and Kastner, 2008). However, it is important to consider the extent to which object-selective areas of dorsal cortex are sensitive to the visual cues that define an object. The shape of an object can be conveyed by pictorial cues such as texture gradients and occlusion, or by stereoscopic cues such as interocular disparity. Unlike ventral cortex, where neural responses to objects are similar irrespective of the cues that define the shape (known as ‘cue invariance’; Grill-Spector et al., 1999), this does not appear to be the case in dorsal cortex (Durand, Celebrini, et al., 2007; Durand, Nelissen, et al., 2007; Sakata et al., 1997; Sary, Vogels, & Orban, 1993). Given the unique selectivity of dorsal areas for object properties (including 3-D shape, absolute size, and egocentric distance), impoverished stimuli such as line drawings and silhouettes (such as those used by Konen and Kastner, 2008) may be inadequate for exploring the underlying selectivity profiles of neural populations in dorsal cortex. Moreover, relying on images of objects to study object representations in dorsal cortex, whether defined by monocular cues or stereoscopic disparity, may not stimulate dorsal cortex as well as solids do, because images do not typically afford genuine motor interaction. Indeed, humans (as well as non-human primates) have evolved to perceive and interact with real objects in real environments, not images (Gibson et al., 1979).

Although image displays permit fine control over stimulus parameters, presentation, and timing, the extent to which they adequately stimulate dorsal cortex, particularly more anterior areas linked with motor preparation and control, is a question that has been surprisingly overlooked in cognitive neuroscience (Gomez et al., 2018; Snow et al., 2011). Although very few studies of dorsal object coding have been conducted with real objects in humans, in the monkey anterior areas such as AIP respond selectively to solids (Murata et al., 2000; Sakata, Taira, Murata, & Mine, 1995). We anticipate that the use of richer, graspable stimuli (i.e., Freud, Culham, Plaut, & Behrmann, 2017; Gomez et al., 2018; Romero, Compton, Yang, & Snow, 2017; Snow et al., 2011; Snow, Skiba, Coleman, & Berryhill, 2014; Squires, Macdonald, Culham, & Snow, 2016) will shed fundamental new insights into the nature of object coding in dorsal cortex.

Taken together, the available evidence argues in favor of unique, veridical object representations in dorsal cortex –representations that support precise goal-directed interactions in an ever-changing dynamic environment. Action-relevant stimulus properties are likely to be determined not only by the nature of the stimulus, but also the current goals and intentions of the observer. Accordingly, object-related responses across the ventral and dorsal streams are likely to be similar in some circumstances, but show striking uniqueness and mutual independence in others, depending on the stimuli and task (Galletti & Fattori, 2018). It will be important in future studies to consider the type of stimuli used and their relevance for genuine goal-directed actions (Gomez et al., 2018; Snow et al., 2011), and which object properties are relevant for the task (Bracci, Daniels & Op de Beeck, 2017). Careful analysis of the nature of representational content, not only in posterior parietal areas (such as IPS1 and IPS2), but also in more anterior regions of dorsal cortex (such as aIPS), are also necessary to develop a more complete understanding of the role of different sub-regions of dorsal cortex in object processing. Ultimately, however, it remains the case that a consideration of the overarching output requirements of the ventral (perception) versus dorsal (action) visual pathways remains a useful and fundamental guiding principle for characterizing the nature of object representations in the brain (Goodale, 2014; Goodale & Milner, 1992, 2018; Milner & Goodale, 1995).

4. Declarations of competing interest

None.

Acknowledgments

This research was supported by a grant to J.C.S. from the National Eye Institute (NEI) of the National Institutes of Health (NIH) under Award Number R01EY026701, a grant to G.E. from the NEI under Award Number F32EY025520, and EPSCoR grants from the National Science Foundation to J.M., J.C.S. and G.P.C. (grant numbers 1632849 and 1632738). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation.

References


Almeida, J., Mahon, B. Z., Zapater-Rabelov, V., Dzuba, A., Cabaco, T., Marques, J. F., & Caramazza, A. (2014). Grasping with the eyes: The role of elongation in visual recognition task has also been shown to predict activity in ventral areas (Sim, Helbig, Graf, & Kiefer, 2015).


