Lessons from human vision for robotic design

Check for
updates

Melvyn A. Goodale[1,2]

Abstract

The visual guidance of goal-directed movements requires transformations of incoming visual information that are different from those required for visual perception. For us to grasp an object successfully, our brain must use just-in-time computations of the object's real-world size and shape, and its orientation and disposition with respect to our hand. These requirements have led to the emergence of dedicated visuomotor modules in the posterior parietal cortex of the human brain (the dorsal visual stream) that are functionally distinct from networks in the occipito-temporal cortex (the ventral visual stream) that mediate our conscious perception of the world. Although the identification and selection of goal objects and an appropriate course of action depends on the perceptual machinery of the ventral stream and associated cognitive modules, the execution of the subsequent goal-directed action is mediated by dedicated online control systems in the dorsal stream and associated motor areas. The dorsal stream allows an observer to reach out and grasp objects with exquisite ease, but by itself, deals only with objects that are visible at the moment the action is being programmed. The ventral stream, however, allows an observer to escape the present and bring to bear information from the past – including information about the function of objects, their intrinsic properties, and their location with reference to other objects in the world. Ultimately then, both streams contribute to the production of goal-directed actions. The principles underlying this division of labour between the dorsal and ventral streams are relevant to the design and implementation of autonomous robotic systems.

Keywords: Perception vs. action, Dorsal visual stream, Ventral visual stream, Tele-assistance, Grasping

1 Introduction

Vision is the most studied and best understood of all our sensory systems. Philosophers have speculated about how vision works for millennia, and over the last two hundred years, scientists have made enormous progress in understanding everything from the intricate organization of the retina to the subtleties of object recognition. The main focus of this enterprise, however, has been directed at explaining how our visual system enables us to perceive the world in all its richness and detail. Far less attention has been paid to how the visual system controls our movement through the world and our interactions with the objects within it. Indeed, until recently, there has been an implicit assumption that the same visual representations that allow us to make sense of the ever-changing patterns of light falling on our retina also provide the information required to control our actions. According to this view, the visual system creates a single "general-purpose" representation of the external world that provides a platform for both cognitive operations as well as the real time control of goal-directed actions. There are good reasons to believe, however, that such a monolithic account is incorrect. In this short review, I argue that the visual guidance of action requires transformations of visual information that are quite different from those required for visual perception. After briefly describing their anatomical substrates, I outline the reasons why vision-for-action has different computational requirements and constraints than vision-for-

Correspondence: mgoodale@uwo.ca
[1]The Brain and Mind Institute, The University of Western Ontario, London, ON N6A 5B7, Canada
[2]Department of Psychology, The University of Western Ontario, London, ON N6A 5C2, Canada

perception. I go on to show how both systems contribute to the production of goal-directed actions, albeit in different but complementary ways. Finally, I make the case that the principles underlying the distinction between the two systems can provide new directions for the design of autonomous and semi-autonomous robots.

2 Two visual systems

Two prominent streams of visual projections arise from primary visual cortex in the human brain [1]: A dorsal visual stream projecting to the posterior parietal cortex and a ventral stream projecting to the occipito-temporal cortex (see Fig. 1). In the early 1990s, David Milner and I proposed a functional account of the division of labor between these two visual streams [2]. According to our two visual systems model, the dorsal stream plays a critical role in the real-time control of action, transforming moment-to-moment information about the location and disposition of objects into the required coordinates for planning and executing actions directed at those objects. The ventral stream, together with its associated cognitive networks, assembles the visual representations of the world that allow us to demarcate objects and events, attach meaning and significance to them, and establish their causal relations. These visual percepts and the associated cognitive operations are critical for building – and accessing – a reservoir of stored knowledge about the world, allowing us to plan future actions and incorporate information from the past into the control of current actions. As one might expect, the two streams are heavily interconnected, reflecting the complementary

Fig. 1 A schematic diagram of the two streams of visual processing in primate cerebral cortex. The retina projects to a number of different brain structures, but illustrated here are three structures that provide input to the cerebral cortex. The ventral stream receives the majority of its visual input from the primary visual cortex (V1), which in turn receives input from the lateral geniculate nucleus (LGNd) of the thalamus. The dorsal stream also receives input from V1, but in addition gets inputs from the superior colliculus (SC) via the pulvinar (Pulv), another nucleus in the thalamus. There are also a small number of projections from the retina directly to the pulvinar, which sends projections to the dorsal stream. The arrows on the inset photograph of the human brain show the approximate route of the two streams within the cerebral hemispheres

roles they play in the production of adaptive behaviour (for a detailed review of the two visual systems model, see refs. [3, 4]). It is also instructive that other sensory pathways in the primate cerebral cortex, such as those associated with somatosensory and auditory processing, also show evidence of a division of labour between object identification and sensory control of movement in space [5, 6].

Well before we put forward the idea of two visual systems, I and others had speculated – on the basis of psychophysical studies in healthy observers – that the visual mechanisms involved in generating our perception of the world are functionally distinct from those mediating the control visually-guided eye and limb movements [7, 8]. But our first clue that this division of labour could be mapped onto the ventral and dorsal visual streams came from studies of the visual deficits and spared visual abilities in neurological patients. One such patient, known by the initials, D.F., suffered damage to her ventral visual stream on both sides of her brain from carbon monoxide poisoning (from a faulty heater). Her dorsal visual stream, however, is relatively intact [9]. Even though her 'low-level' visual abilities are largely within the normal range, D.F. can no longer recognize everyday objects or the faces of her family and friends. Even now, decades after her accident, her deficit remains so profound that she cannot discriminate between simple geometric shapes, such as a triangle and square. Nevertheless, she has no difficulty describing an object's colour or visual texture (whether it has a glossy or matte finish, for example) – and can readily tell from visual inspection if an object is made of metal, wood, cloth, or some other material [10]. It is the shape of the object that she has problems with. It should be emphasized that she has no trouble identifying the shape of familiar objects by touch. Her deficit in form recognition of such objects is entirely restricted to vision. Moreover, her failure to identify an object is not due to a disconnection between the visual percept of an object and associated semantic information about that object. When D.F. is asked to copy a line drawing, for example, her renditions bear almost no relationship to the depicted object [11], even though she can draw reasonably well from memory or on the basis of haptic exploration of an object. In short, she appears to have a fundamental deficit in the ability to *perceive* the shapes of objects. Historically, neurologists have referred to this deficit as *apperceptive agnosia* [12], or more recently, as *visual form agnosia* [13]. The term 'agnosia' comes from the ancient Greek and means 'ignorance' or 'not knowing'.

What is truly remarkable about D.F., however, is that despite her visual form agnosia, she shows strikingly
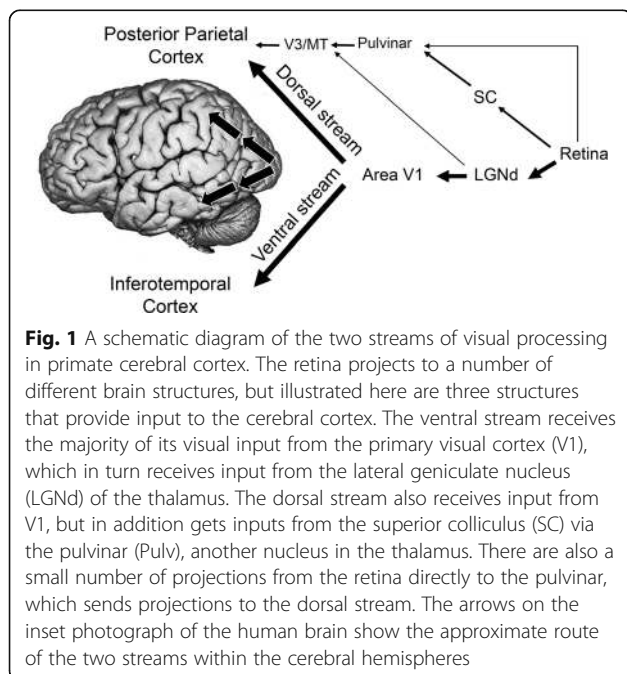
accurate guidance of her hand movements when she attempts to pick up the very objects she cannot identify. Thus, as Fig. 2 shows, when she reaches out to grasp objects of different sizes, the opening of her hand is scaled mid-flight to the width of the object, just as it is in people with normal vision [9]. Similarly, she rotates her hand and wrist quite normally when she reaches out to grasp objects that are positioned in different orientations with respect to her hand [14]. She also avoids colliding with other objects in the workspace as her hand moves towards the goal object [15]. Moreover, when confronted with objects of different shapes, she places her fingers on stable grasp points on the surface of each object, so that the applied forces pass through the object's centre of gravity [16]. Yet she is unable to distinguish amongst these objects when they are presented to her in simple discrimination tests. She even fails in manual 'matching' tasks in which she is asked to show how wide an object

Fig. 2 Grasping an object vs. manually estimating its size. The photographs illustrate a participant reaching out and grasping a rectangular object using a precision grip (left) or manually estimating its width with the same finger and thumb (right). Both movements are typically recorded with an optoelectronic motion tracking system. Maximum grip aperture of the grasping hand is achieved approximately 70% of the way towards the goal object. The hand opens wider than the width of the object, even though the opening is scaled to objects of different widths. The two graphs show the size of the aperture between the index finger and thumb during object-directed grasping and manual estimates of object width for D.F., a patient with a bilateral ventral-stream lesion, and RV, a patient with a bilateral dorsal-stream lesion. DF showed excellent grip scaling, opening her hand wider for the 50-mm-wide object than for the 25-mm-wide object (individual trials marked as open diamonds). D.F.'s manual estimates of the width of the two objects, however, were grossly inaccurate and showed enormous variability from trial to trial. RV was able to indicate the size of the objects reasonably well, but her maximum grip aperture in flight was not well-tuned. She simply opened her hand as wide as possible on every trial

is by opening her index finger and thumb a corresponding amount (see Fig. 2).

The presence of spared visual control of grasping in D.F., who has profound deficits in the perception of object shape, lends considerable support to the idea that vision-for-action depends on mechanisms that are quite separate from those involved in vision-for-perception. In addition, the fact that D.F. and other patients with visual form agnosia [17] have damage in their ventral stream suggests that this particular stream plays a critical role in visual perception – and that the intact dorsal stream in these patients may be mediating their spared visuomotor skills. This conclusion is borne out by observations in patients who have dorsal-stream damage (in the posterior parietal cortex) but an intact ventral stream. These patients typically show striking deficits in their ability to reach out and grasp objects properly even though they can accurately describe the objects' location, orientation, shape, and size (see Fig. 2) [18–20]. In addition, they have difficulty avoiding obstacles in the workspace of their hand [21]. These deficits are not due to some sort of basic motor problem. The patients often have no difficulty reaching out and touching different parts of their body on command, for example, moving their hand quickly to locations on their body that are touched by the examining physician or experimenter. Their deficit is truly visuomotor in nature. Neurologists refer to these deficits in visually guided reaching and grasping after damage to the dorsal stream as *optic ataxia*, where 'ataxia' is derived from medical Latin, ultimately from Ancient Greek, and means 'lack of order' or 'disorder'.

In summary then, it was this pattern of deficits and spared visual abilities observed in individuals with damage to the either the ventral or dorsal stream, coupled with additional evidence from neurophysiological and behavioural studies in non-human primates, that first led us to propose that the division of labour between vision-for-perception and vision-for-action could be mapped on to the ventral and dorsal streams respectively. More recent evidence from neuroimaging in patients and neurologically healthy participants has provided additional and convincing support for this proposal (for review, see refs. [17, 22]).

3 Different metrics and frames of reference for perception and action

Although the evidence is compelling for the idea that the ventral and dorsal streams play different but complementary roles in our visual lives, the question arises as to why this division of labour evolved. After all, why couldn't one general-purpose visual system do the job? The answer to this question lies in differences in the nature of the transformations on the incoming visual
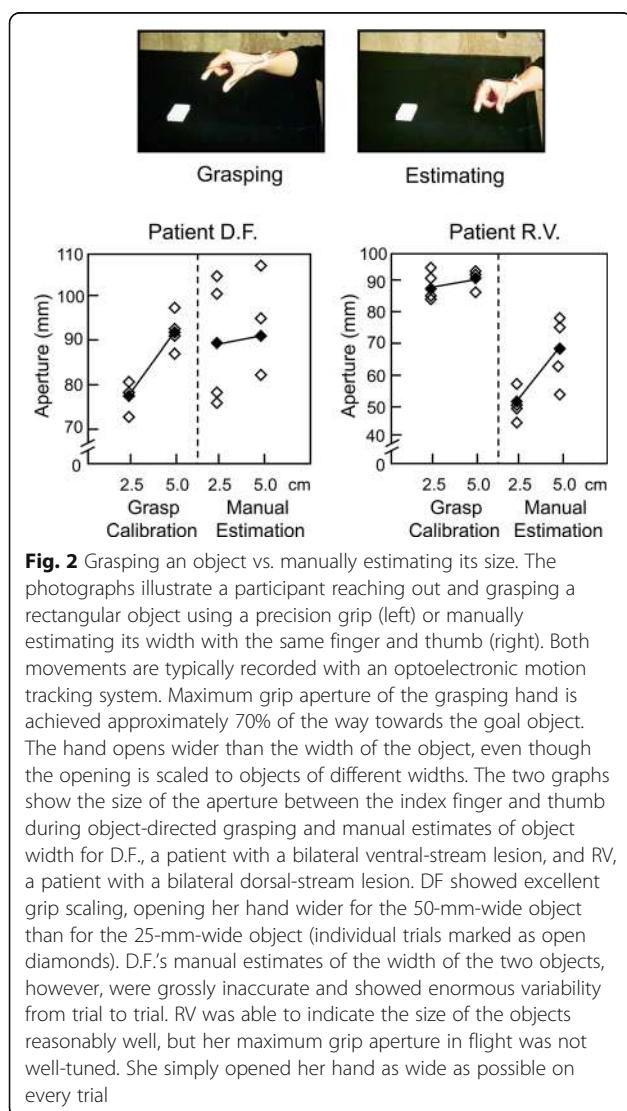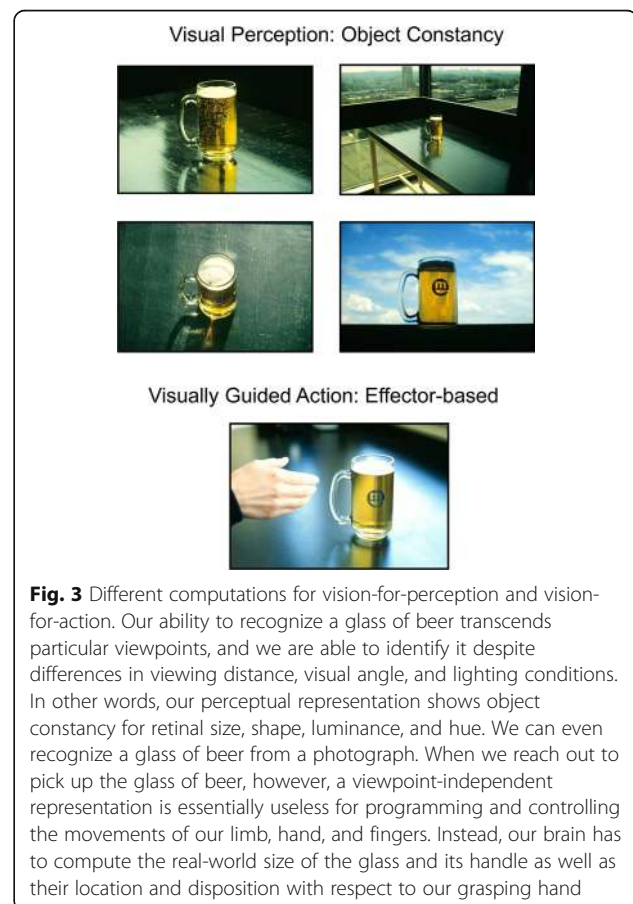
information that are required for action vs. perception. To be able to grasp a glass of beer successfully, for example, it is essential that the brain compute the actual size of the beer glass and its orientation and position with respect to the hand we intend to use to pick it up (Fig. 3). There are also critical temporal constraints as well. The location and disposition of a target object with respect to the one's hand, for example, can change radically from one moment to the next. As a consequence, the required motor coordinates for an action have to be calculated at the very moment the movement occurs, a rapid just-in-time computation. All of this demands that the neural circuitry supporting such rapid conversion of visual information into action be intimately interconnected with motor systems in the cerebral cortex, midbrain, brainstem, and cerebellum. As it happens, the dorsal stream fulfils these requirements, having extensive reciprocal connections with premotor cortex and projections to the superior colliculus and the dorsolateral pontine nuclei in the brainstem (which are connected to the cerebellum). In short, the dorsal stream is well poised for the visual control of action. It is worth emphasizing that the superior colliculus and other sub-cortical structures receive direct input from the retina and are capable of initiating visually guided actions, such as saccadic eye movements, entirely on their own. One important way that the dorsal visual stream exercises control over visually guided movements is to modulate the activity of these sub-cortical structures. For more details about the sub-cortical recipients of retinal inputs, see ref. [3].

Perceiving the world presents our brain with quite a different problem. Although computing the real-world size of a glass of beer and its location and disposition with respect to our hand at a particular moment in time is essential to picking it up successfully, recognizing it as a glass of beer requires that our perceptual processing transcend the differences in viewpoint that typically occur from one occasion to the next (Fig. 3). In other words, we need to be able to recognize an object despite dramatic changes in the projection of its image onto our retina. Moreover, it is important for stable perception that we encode the size, orientation, and location of objects relative to each other. Such a scene-based frame of reference preserves information about spatial relationships amongst objects (as well as their relative size and orientation) as we move around, information that is critical for understanding the world. By working with perceptual representations that are scene-based, we are able to maintain the constancies of size, shape, color, lightness, and relative location, over time and across different viewing conditions. Although there is much debate about the way in which this information is computed and encoded, it is clear that it is the identity of the object and its location within the scene, not its exact

disposition with respect to the observer that is of primary concern to the perceptual system. In fact, the perceptual networks in the ventral stream enable us to watch and make perfect sense of movies playing on television or in the theater, even though the objects on the screen bear no relationship to their real-world size, and the events that are depicted are not unfolding within an egocentric frame of reference. It is the cinematographer, not us, who is in charge of showing us a particular scene and selecting the viewpoint we are presented with. Yet we have no trouble at all in understanding what is happening, provided certain conventions about camera angles are observed. What makes this possible of course is that our perception of the world relies almost entirely on relational metrics and scene-based frames of reference. As a vehicle for the control of action, however, movies are essentially hopeless.

Importantly too, perceptual representations of objects are available over a much longer time scale than the just-in-time computations that drive visually guided actions. We recognize objects we have seen minutes, hours, days – or even years before. Thus, we might recognize a person walking down the street as someone we met several months ago – presumably because our

Fig. 3 Different computations for vision-for-perception and vision-for-action. Our ability to recognize a glass of beer transcends particular viewpoints, and we are able to identify it despite differences in viewing distance, visual angle, and lighting conditions. In other words, our perceptual representation shows object constancy for retinal size, shape, luminance, and hue. We can even recognize a glass of beer from a photograph. When we reach out to pick up the glass of beer, however, a viewpoint-independent representation is essentially useless for programming and controlling the movements of our limb, hand, and fingers. Instead, our brain has to compute the real-world size of the glass and its handle as well as their location and disposition with respect to our grasping hand

current percept is compared to a stored representation of that person. But when we reach out to shake that person's hand, the visuomotor circuits driving the movements of our arm and hand do not refer to stored coordinates from the past, i.e., from the last time we shook hands. Instead, the required coordinates are computed on the spot by visuomotor mechanisms in the dorsal stream. Interestingly, a recent study using recurrent neural networks showed that 'ventral-stream' object classification requires longer memory than 'dorsal-stream' orientation classification [23].

Visual perception allows us to make sense of the world, plan a vast range of different courses of action with respect to the objects and events we have identified, and share our thoughts and plans with others. The perceptual mechanisms that allow our brains to do this need not be linked directly to specific motor outputs, but instead are likely to access these outputs via cognitive systems involved in memory, semantics, spatial reasoning, planning, and communication. In other words, there are likely to be a lot of cognitive buffers between perceiving the world and acting on it. It is perhaps not surprising therefore that the ventral visual stream, which mediates our perception of the world, has few or no direct connections with motor circuits either in the cerebral cortex, midbrain, or the brainstem, but instead is intimately interconnected with areas in the temporal lobe and prefrontal cortex involved in memory, decision-making, language, and social behaviour [24].
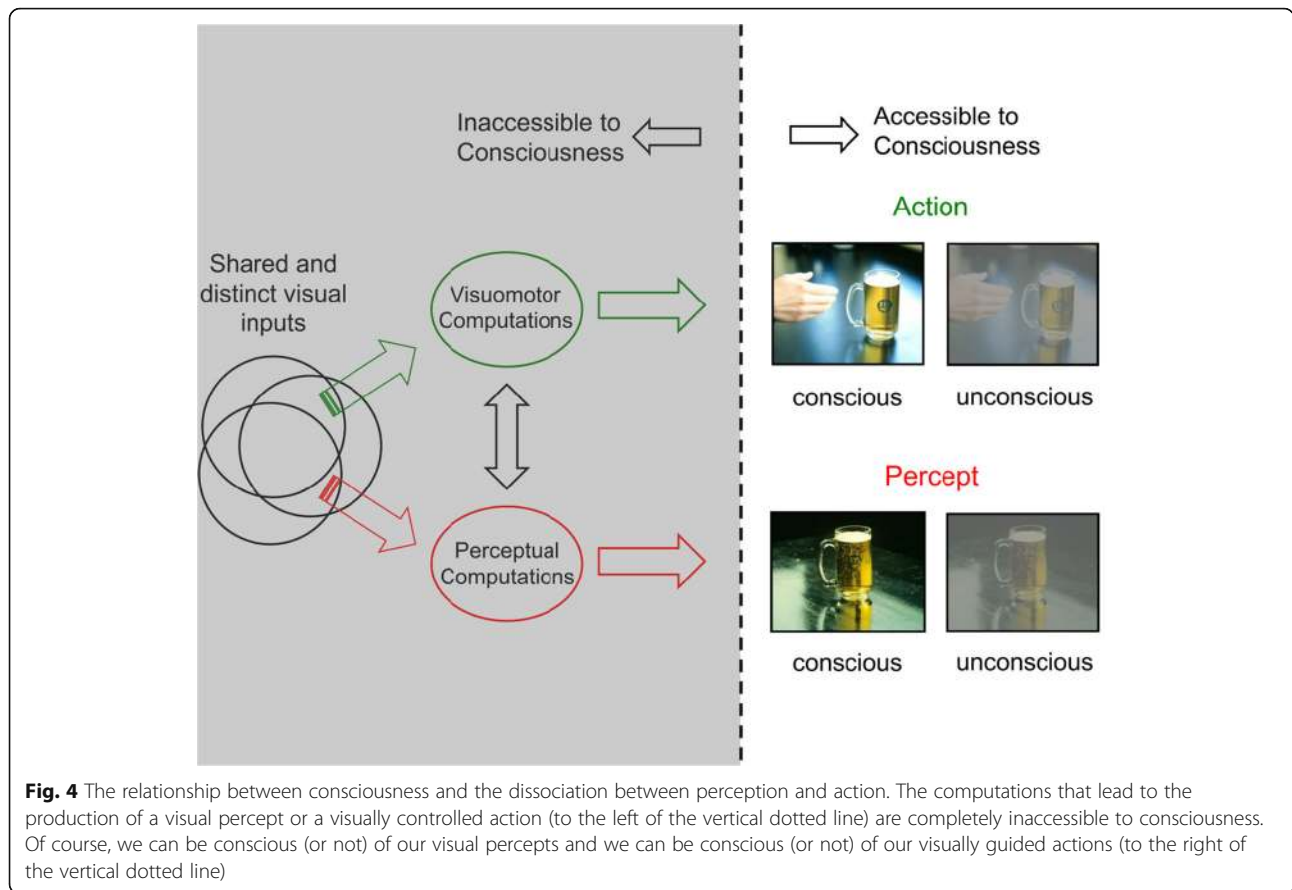
4 The relationship to consciousness

The perceptual mechanisms in the ventral stream give rise to conscious visual percepts. We *experience* a world beyond our bodies. Consciousness is a hotly debated topic in both philosophy and cognitive neuroscience, but without delving into the contentious issues surrounding the nature of qualia, most of us will agree that there is little debate about the fact we can describe the objects and events we see when we look out at the world. Of course, we are not always conscious of every percept created by the ventral visual pathway, even though those unseen objects and events can influence our subsequent behaviour [25]. But what about the dorsal stream? Here I would argue that we are often conscious (but not always) of the actions that are programmed and controlled by visuomotor mechanisms in the dorsal stream. In other words, we have a sense of agency when we perform many visually guided actions [26]. But importantly, we are not conscious and never can be of the visual information that contributes to the computations carried out by those mechanisms. Of course, by the same token, we are not privy to the visual information used by the ventral stream to construct our

percepts of the world, even though those percepts are 'visual' in nature. As Fig. 4 summarizes, we can be conscious (or not) of the products of the two streams – a visual percept in the case of the ventral stream, and an action in the case of the dorsal stream – but we can never be conscious of the visual information that contributed to the construction of a percept or the performance of an action. It should be noted as well that our ventral visual stream allows us to perceive our hand moving towards our glass of beer, but it plays no causal role in the real-time control of that movement [27]. Although we might believe it is our percept of the glass that provides the real-time control of grasp, that is simply an illusion, what the philosopher, Andy Clark, has called the "assumption of experience-based control" [28].

5 Biological tele-assistance

Clearly, the ventral and dorsal streams must work closely together in the generation of purposive behavior. How might this occur? A productive interaction between the two streams would undoubtedly take advantage of the complementary differences in their computational constraints. A metaphor from engineering that captures the nature of that interaction is *tele-assistance*, a robotic control system whereby a human operator, who has identified a goal object and decided what to do with it, communicates with a semi-autonomous robot that actually performs the required motor act on the flagged goal object, in what is typically a dangerous or otherwise inaccessible environment [29, 30]. The robot itself makes use of its onboard range-finders and instruments to deal with flagged object. In terms of this tele-assistance metaphor, the perceptual networks in the ventral stream via their links with other cognitive systems would be the human operator. These networks identify a relevant goal object in a scene and select an appropriate course of action to deal with that object. Once a particular goal object, such as a glass of beer, has been flagged (presumably by means of an attentional process), semi-autonomous "robotic" systems in the dorsal stream (in conjunction with related circuits in premotor cortex, basal ganglia, midbrain, and brainstem) would then perform the just-in-time computations required to transform visual information about the glass into the appropriate coordinates for the desired motor act. Of course, in other situations, where visual stimuli, such as an obstacle, are particularly salient, the visuomotor mechanisms in the dorsal stream will operate without any immediate supervision by ventral stream perceptual mechanisms (as would be the case with a semi-autonomous robot). Similarly, once the dorsal stream networks have locked on to the target, any subsequent movement of the target will be tracked automatically,

Fig. 4 The relationship between consciousness and the dissociation between perception and action. The computations that lead to the production of a visual percept or a visually controlled action (to the left of the vertical dotted line) are completely inaccessible to consciousness. Of course, we can be conscious (or not) of our visual percepts and we can be conscious (or not) of our visually guided actions (to the right of the vertical dotted line)

much like what typically occurs if a robot (or perhaps a drone or missile) is pursuing a moving target.

One has to be careful, of course, not to push the tele-assistance metaphor too far. For one thing, the ventral stream by itself cannot be construed as an intelligent operator that can make assessments and plans. Some sort of top-down executive control is required that almost certainly engages high-level cognitive mechanisms in the frontal lobe that initiate the operation of attentional search and thus set the whole process of planning and goal selection in motion (for review, see refs. [31, 32]). There is also the problem of how the ventral and dorsal streams communicate information about the goal object that has been identified. There are complex interconnections between the ventral stream, frontal lobe networks, and the dorsal stream that undoubtedly play a role in modulating the activity of specialized networks in premotor cortex and the dorsal stream that mediate the control of voluntary eye movements, as well as covert shifts of attention [33, 34]. In terms of the tele-assistance metaphor, these eye-movement circuits can be seen as acting like the videocam on a robot that the operator uses to scan the visual scene to search of possible goal objects. For a more detailed discussion of the interactions between the two streams and their related networks elsewhere in the brain, see ref. [17].

Although the tele-assistance model is somewhat fanciful, it does provide a useful engineering framework for understanding how the functions of the two streams complement one another. In fact, the fidelity of the analogy underscores once again the fact that the evolution of the mammalian visual system has anticipated modern developments in robot engineering and machine vision by several million years. Of course in drawing the analogy with tele-assistance, one should not underestimate future developments in the design of autonomous robots. Clearly, engineers are making enormous progress on this front, and it is likely that the role of the human operator will eventually be incorporated into the design of intelligent autonomous robots. But one can speculate about the kind of visual system such a robot might have. The lessons learned from biology suggest that there would be little prospect of success in trying to give such a robot a general-purpose visual system, one that both recognizes objects in the world and guides the robot's movements. As I have argued thus far, the computational demands of scene analysis and object recognition and are simply incompatible with the computational demands of visuomotor control. Indeed, a much more effective design for the visual system in an intelligent autonomous robot would be to emulate the division of

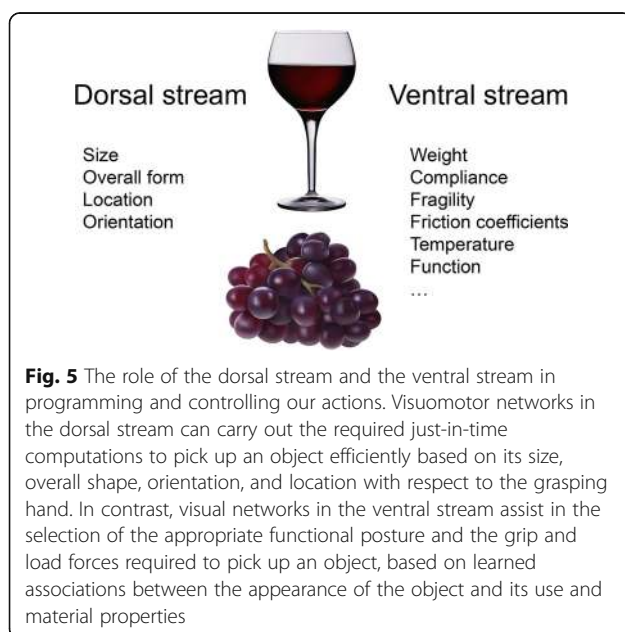labor between the ventral and dorsal visual streams in the primate brain.

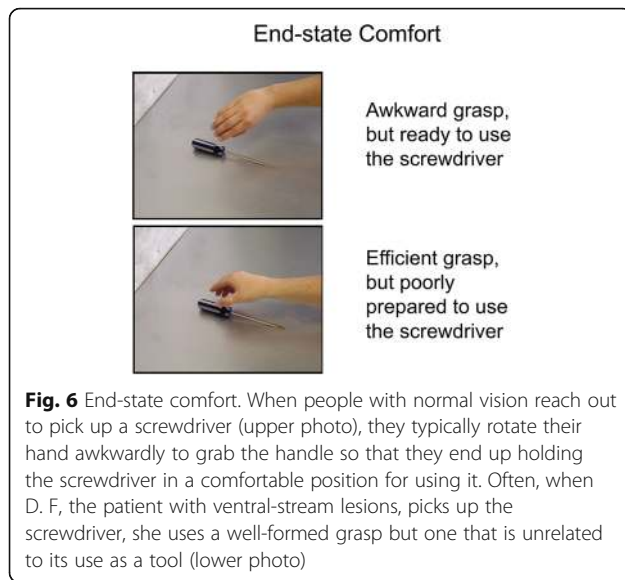## 6 The contributions of the ventral stream to action

It is important to acknowledge that even though the ventral visual stream has no direct connections with motor networks, it still has a profound effect on certain aspects of motor programming, notably those that depend on information that cannot be derived directly from the projected image on retina. Although the visuomotor networks in the dorsal stream are quite capable of using this kind of bottom-up information to compute size, shape, location and orientation of when we reach out to pick up a beer glass, for example, they cannot compute the required grip and load forces that would need to be applied the moment contact is made with the surface of the glass, well before any feedback from touch and other somatosensory receptors comes into play. Those forces must be scaled appropriately for an object's mass, compliance, surface friction, and (in the case of a glass of beer) how much beer is in the glass – information that can be gleaned only through experience (Fig. 5). This is where the ventral stream comes into play. The perceptual mechanisms in the ventral stream enable us to identify the features of the object and access stored information about the required grip and load forces that was acquired through past interactions with that object and/or similar objects. For example, by virtue of your experience, you would automatically apply much greater force to pick up a large rock than you would to pick up a piece of polystyrene of the same size unless of

Fig. 5 The role of the dorsal stream and the ventral stream in programming and controlling our actions. Visuomotor networks in the dorsal stream can carry out the required just-in-time computations to pick up an object efficiently based on its size, overall shape, orientation, and location with respect to the grasping hand. In contrast, visual networks in the ventral stream assist in the selection of the appropriate functional posture and the grip and load forces required to pick up an object, based on learned associations between the appearance of the object and its use and material properties

course the polystyrene had been painted to look like a stone, as it might on a film set (and that would be quite a surprise). Recent neuroimaging studies have shown that, when people pick up objects of different materials, such as wood or metal, activity increases just before the object is grasped in areas of the ventral stream that process the visual texture of those objects [35]. Of course, bigger objects made of the same material weigh more than smaller ones – and thus more force is required to pick them up. Theoretically, size in this case could be computed by the dorsal stream, since it already has been shown to compute size to scale the aperture of the grasping hand. But behavioural studies have shown that the computation of size for the anticipated application of the forces required to lift an object are almost certainly carried in the ventral not the dorsal stream. As a consequence, size computations in the ventral stream are more likely to be subject to contextual effects (e.g., the size of nearby objects) than are size computations in the dorsal stream [17, 36, 37].

Another critical contribution of the ventral stream to our interactions with objects can be seen in the use of tools. In order to pick up a tool properly so that it is ready to use, we must first recognize it and select what part of the tool to grasp. We typically pick up a knife, for example, by grasping its handle not its blade. The perceptual mechanisms in the ventral stream (along with the cognitive modules with which they are connected) allow us to recognize the knife, select its handle as our goal, and choose the appropriate functional hand posture. But it is the dorsal stream, of course, that specifies the parameters of the grasping movement based on the width and shape of the handle as well as its location and disposition with respect to our hand at the moment we initiate the action. Both streams work together then when we interact with tools.

The complementary contributions of the two visual streams can be seen in the way we deal with a tool, such as a screwdriver for example, when the handle is pointed directly away from us rather than towards us. When that occurs, we typically turn our hand right around in a somewhat awkward fashion and grasp it by the handle so that it is ready to use (see Fig. 6). Not only does the ventral stream enable us to recognize the screwdriver but the hand posture we adopt when grasping it is determined by our intentions. Take the case of a glass of beer. If we intend to drink the beer from the glass, we grasp it so that we can easily convey the glass to our lips – but if we intend instead to put an empty glass in the dishwasher, we grasp it in quite the opposite way, so that we can easily rotate it and place it top down on the rack of the dishwasher. We select our grasp posture to ensure that we achieve what is sometimes called 'end-state

Fig. 6 End-state comfort. When people with normal vision reach out to pick up a screwdriver (upper photo), they typically rotate their hand awkwardly to grab the handle so that they end up holding the screwdriver in a comfortable position for using it. Often, when D. F, the patient with ventral-stream lesions, picks up the screwdriver, she uses a well-formed grasp but one that is unrelated to its use as a tool (lower photo)

comfort' [38]. The ability to achieve end-state comfort depends not only on perceptual processing in the ventral stream but also on accessing stored information about how the hand should grip the object in order to achieve end-state comfort. But again the specification of the required parameters for grasping the object, even when the selected hand posture is awkward, depends on just-in-time computations by the visuomotor networks in the dorsal stream.

If it is the ventral stream that enables us to identify the function of an object (and select the appropriate hand posture by virtue of its links with stored information about different functional hand postures), then patient D.F., who has a damaged ventral stream, should be expected to treat a manufactured tool in much the same way as she would a stick. For example, when asked to pick up a screwdriver with the handle pointed away from her (without telling her what it is), she should not rotate her hand awkwardly as people with normal vision would do, making sure that they have hold of the handle in anticipation of achieving end-state comfort, but should instead grasp it using an efficient but inappropriate hand posture. In fact, that is exactly what she does [14]. Her grasp is perfectly matched to the screwdriver's size, shape, and location, but shows no indication that she understands its function; i.e., she often ends up grasping the shaft of the screwdriver, rather than its handle – and only then does she rotate it in her hand so that she can hold it properly. It seems that her (unsupervised) dorsal stream is still working well, even though the damage to her ventral stream prevents her from identifying what the object is that she picking up.

Recent neuroimaging studies in the human brain have revealed a complex network of areas that are involved in tool use [39]. Circuits in both the ventral and the dorsal

stream figure prominently in this network. The interconnectivity between these circuits changes dynamically with changes in task and visual input. Thus, when people look at images of tools (as opposed to graspable non-tools) while their brains are being scanned, there is an increase in the reciprocal connectivity between the ventral and the dorsal visual stream [40].

The computations carried out by the visuomotor networks in the dorsal stream are by design stuck in the present. In other words, they carry out the just-in-time computations required to ensure that actions are directed to the right place and reflect the size and disposition of the goal object at that particular instant in time. In contrast, the ventral stream enables us to make use of previously learned information about goal objects, such as information about their material properties, their functions, and their location with reference to other objects in the world. Both streams, it seems, contribute to the production of goal-directed actions, but in quite different ways.

7 Implications for the design of autonomous robots

These interactions between visual processing in the ventral and dorsal streams of the human brain could eventually be emulated in the design of autonomous robots [41, 42]. Ventral-stream-like modules in such a robot, working in concert with stored information about the world (both acquired and pre-determined) and engaging cognitive modules capable of decision-making, planning, and communication, could learn to parse a scene and to recognize objects and the function of those objects. Information about those objects could then be used to model the potential outcome of different courses of action, allowing the robot to select appropriate actions for interacting with objects efficiently to achieve the desired goal. The actual programming of the action, however, would be guided by dorsal-stream-like circuits that carry out the necessary just-in-time computations for specifying the kinematics of the action on the basis of visual input about the size, shape, location and disposition of the goal object with respect to the robot's effectors. It is worth remembering that biology and biological principles have inspired the design of successful engineering creations from aircraft to artificial heart valves [43, 44]. The development of a new generation of intelligent autonomous robots is likely to be no difference. Already, some theorists and engineers are incorporating elements of the duplex visual system of the primate brain and associated cognitive modules into the design of robots capable of sophisticated grasping [45–49]. But even without explicitly bio-mimetic approaches to the design of such robots, the eventual implementation of the visual modules and their interactions would almost certainly

converge on a functional architecture that resembles the neural networks that have evolved over millennia in the primate brain.

Code availability
Not applicable.

Author information
Although he was born in England, Prof. Goodale was educated entirely in Canada. After completing his Ph.D. in Psychology at Western University, he returned to England where he worked as a Postdoctoral Fellow at Oxford University with Professor Larry Weiskrantz. After two years at Oxford, Prof. Goodale accepted a position in the School of Psychology at the University of St. Andrews in Scotland. After six years at St. Andrews, he returned to Canada, where he now holds the Canada Research Chair in Visual Neuroscience and is a Distinguished University Professor in the Department of Psychology and the Department of Physiology and Pharmacology. He is the Founding Director of Western's Brain and Mind Institute.

His early work, in which he demonstrated that the visual control of action is functionally independent of conscious visual perception, laid the foundation for the influential Two-Visual-Systems account of high-level vision. This account provides a convincing resolution to conflicting accounts of visual function that have characterized much of the work in the field for the last one hundred years. Over the last two decades, he has carried out neuropsychological, psychophysical, and neuroimaging research that has refined and extended the Two-Visual-Systems proposal. This account of the functional organization of vision and its underlying neural substrates has had an enormous influence in the life sciences and medicine, and is now part of almost every textbook in vision, cognitive neuroscience, and psychology. In addition, his ideas have also influenced researchers working in machine vision and robotic control, as well as philosophers interested in consciousness. In addition to his research activities, Prof. Goodale play a pivotal role in establishing the graduate program in Neuroscience at Western Ontario, for which he was awarded Western's E.G. Pleva Award for Contributions to Teaching in 1994. He serves on the editorial board of a number of academic journals, including Proceedings of the Royal Society B and Neuropsychologia. He was the Inaugural President of the Canadian Society for Brain, Behaviour, and Cognitive Neuroscience (CSBBCS) and President of the Association for the Scientific Study of Consciousness.

In 1999, Prof. Goodale was awarded the Donald O. Hebb Distinguished Contribution Award (CSBBCS). In 2007, he was awarded the Hellmuth Prize for Scientific Achievement (Western) and, in 2008, the Richard C. Tees Award for Distinguished Leadership (CSBBCS). He is a Fellow of the Royal Society of Canada and the Royal Society of London (UK). In 2016, he was appointed as the Ivey Fellow by the Canadian Institute for Advanced Research (CIFAR).

Author's contributions
Melvyn A. Goodale wrote this review. The author read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Competing interests
None.

References
1. L.G. Ungerleider, M. Mishkin, in *Mansfield*, ed. by D. J. Ingle, M. A. Goodale, RJW. Analysis of Visual Behavior (MIT Press, Cambridge, 1982), p. 549
2. M.A. Goodale, A.D. Milner, Trends. Neurosci. **15**, 20 (1992)
3. A.D. Milner, M.A. Goodale, *The Visual Brain in Action*, 2nd edn. (OUP, Oxford UK, 2006), pp. 1–297
4. M.A. Goodale, A.D. Milner, *Sight Unseen: An Exploration of Conscious and Unconscious Vision*, 2nd edn. (OUP, Oxford UK, 2013), pp. 1–218
5. H.C. Dijkerman, E.H. De Haan, Behav. Brain. Sci. **30**, 189 (2007)
6. J.P. Rauschecker, Eur. J. Neurosci. **41**, 579 (2015)
7. B. Bridgeman, S. Lewis, G. Heit, M. Nagle, J. Exp. Psychol. [Hum. Percept.] **5**, 692 (1979)
8. M.A. Goodale, D. Pélisson, C. Prablanc, Nature **320**, 748 (1986)
9. M.A. Goodale, A.D. Milner, L.S. Jakobson, D.P. Carey, Nature **349**, 154 (1991)
10. G.K. Humphrey, M.A. Goodale, L.S. Jakobson, P. Servos, Perception **23**, 1457 (1994)
11. P. Servos, M.A. Goodale, G.K. Humphrey, Neuropsychologia **31**, 251 (1993)
12. H. Lissauer, Arch. Psychiatr. **21**, 222 (1890)
13. D.F. Benson, J.P. Greenberg, Arch. Neurol. **20**, 82 (1969)
14. D.P. Carey, M. Harvey, A.D. Milner, Neuropsychologia **34**, 329 (1996)
15. N.J. Rice, R.D. McIntosh, I. Schindler, M. Mon-Williams, J. Demonet, A.D. Milner, Intact automatic avoidance of obstacles in patients with visual form agnosia. Exp. Brain Res. **174**, 176 (2006)
16. M.A. Goodale, J.P. Meenan, H.H. Bülthoff, D.A. Nicolle, K.J. Murphy, C. Racicot, Curr. Biol. **4**, 604 (1994)
17. M.A. Goodale, Vis. Res. **51**, 1567 (2011)
18. R. Bálint, Monatsschr. Psychiatr. Neurol. **25**, 51 (1909)
19. M.-T. Perenin, A. Vighetto, Brain **111**, 643 (1988)
20. L.S. Jakobson, Y.M. Archibald, D.P. Carey, M.A. Goodale, Neuropsychologia **29**, 803 (1991)
21. I. Schindler, N.J. Rice, R.D. McIntosh, Y. Rossetti, A. Vighetto, A.D. Milner, Automatic avoidance of obstacles is a dorsal stream function: Evidence from optic ataxia. Nature Neurosci. **7**, 779 (2004)
22. J.P. Gallivan, M.A. Goodale, in *Handbook of Clinical Neurology, Vol. 151: The Parietal Lobe*, ed. by G. Vallar, H. P. Coslett. (Elsevier, Amsterdam, 2018), p. 449
23. A. Alipour, J. Beggs, J. Brown, T. James, bioRxiv 2020.09.30.321299. https://doi.org/10.1101/2020.09.30.321299
24. M. Glickstein, Trends. Neurosci. **23**, 613 (2000)
25. S. Kouider, S. Dehaene, Philos. Trans. R. Soc. **362**, 857 (2007)
26. P. Haggard, Nat. Rev. Neurosci. **18**, 196 (2017)
27. A.D. Milner, Proc. R. Soc. B **279**, 2289 (2012)
28. A. Clark, Philos. Rev. **110**, 495 (2001)
29. P.K. Pook, D.H. Ballard, Rob. Auton. Syst. **18**, 259 (1996)
30. M.A. Goodale, G.K. Humphrey, Cognition **67**, 181 (1998)
31. R. Desimone, J. Duncan, Ann. Rev. Neurosci. **18**, 193 (1995)
32. M.A. Goodale, A.M. Haffenden, Adv. Neurol. **93**, 249 (2003)
33. M. Corbetta, J.M. Kincade, G.L. Shulman, J. Cogn. Neurosci. **14**, 508 (2002)
34. J.W. Bisley, M.E. Goldberg, Adv. Neurol. **93**, 141 (2003)
35. J.P. Gallivan, J.S. Cant, M.A. Goodale, J.R. Flanagan, Curr. Biol. **24**, 1866 (2014)
36. S.R. Jackson, A. Shaw, J. Exp. Psychol. Hum. Percept. Perform. **26**, 418 (2000)
37. M.A. Goodale, T. Ganel, in *Oxford Handbook of Perceptual Organization*, ed. by B. J. Wagemans. (OUP, Oxford, 2015), p. 672
38. D.A. Rosenbaum, F. Marchak, H.J. Barnes, J. Vaughan, J.D. Slotta, M.J. Jorgensen, MJ, in *Attention and Performance XIII: Motor Representation and Control*, ed. by M. Jeannerod. (Erlbaum, Hillsdale, NJ, 1990), p. 321
39. J.P. Gallivan, D.A. McLean, K.F. Valyear, J.C. Culham, Elife **2**, e00425 (2013)
40. J. Chen, J.C. Snow, J.C. Culham, M.A. Goodale, What role does "elongation" play in "tool-specific" activation and connectivity in the dorsal and ventral visual streams? Cereb. Cortex **28**, 1117 (2018)
41. E. Chinellato, A.P. Del Pobil, *International Work-Conference on the Interplay between Natural and Artificial Computation* (Springer, Heidelberg, 2005) p. 366
42. E. Chinellato, A.P. Del Pobil, *Visual Neuroscience of Robotic Grasping: Achieving Sensorimotor Skills through Dorsal-Ventral Stream Integration* (Springer, Heidelberg, 2016), pp. 1–165

43. M. Helms, S.S. Vattam, A.K. Goel, Design Stud. **30**, 606 (2009)
44. L.H. Shu, K. Ueda, I. Chiu, H. Cheong, CIRP Ann. Manuf. Technol. **60**, 673 (2011)
45. A. Saxena, J. Driemeyer, A.Y. Ng, Int. J. Robot. Res. **27**, 157 (2008)
46. L.Y. Ku, L.E. Learned-Miller, R. Grupen, *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2017), p. 2434
47. R. Detry, J. Papon, L. Matthies, *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2017), p. 3266
48. E. Jang, S. Vijayanarasimhan, P. Pastor, J. Ibarz, S. Levine, Proc. Mach. Learn. Res. **78**, 119 (2017)
49. G. Ferretti, E. Chinellato, in *Blended Cognition*, ed. by B. J. Vallverdú, V. Müller. (Springer, Cham, 2019), p. 99

Publisher's Note