



Conscious Vision in Action

Robert Briscoe,^a John Schwenkler^b

^a*Department of Philosophy, Ohio University*

^b*Department of Philosophy, Florida State University*

Received 10 September 2012; received in revised form 27 August 2014; accepted 29 August 2014

Abstract

It is natural to assume that the fine-grained and highly accurate spatial information present in visual experience is often used to guide our bodily actions. Yet this assumption has been challenged by proponents of the Two Visual Systems Hypothesis (TVSH), according to which visuo-motor programming is the responsibility of a “zombie” processing stream whose sources of bottom-up spatial information are entirely non-conscious (Clark, 2007, 2009; Goodale & Milner, 1992, 2004a; Milner & Goodale, 1995/2006, 2008). In many formulations of TVSH, the role of conscious vision in action is limited to “recognizing objects, selecting targets for action, and determining what kinds of action, broadly speaking, to perform” (Clark, 2007, p. 570). Our aim in this study is to show that the available evidence not only fails to support this dichotomous view but actually reveals a significant role for conscious vision in motor programming, especially for actions that require deliberate attention.

Keywords: Consciousness; Vision; Motor control; Dual Visual Systems; Attention; Dorsal stream; Ventral stream

1. Introduction

Visual experiences represent the spatial properties of visible objects and surfaces, including their distances, directions, orientations, shapes, and sizes. At close range and under ecologically normal viewing conditions, conscious perception of spatial layout is highly accurate. While perceived distances are significantly foreshortened for objects located more than 30 m away, in what Cutting and Vishton (1995) call *vista space*, the egocentric distances of objects located up to 2 m away, in *personal space*, are perceived with nearly metric accuracy. The difference in precision is a function of the spatial information available to the visual system. Estimates of depth for objects in personal

Correspondence should be sent to John Schwenkler, Department of Philosophy, 151 Dodd Hall, Florida State University, Tallahassee, FL 32306-1500. E-mail: jlschwenkler@gmail.com

space are powerfully constrained by stereopsis, convergence, and accommodation. As distances lengthen, these binocular sources of depth information drop off in effectiveness, and the visual system must rely increasingly on somewhat less precise monocular or “pictorial” cues in the light sampled by the eyes.

Given the accuracy with which the spatial properties of objects in personal space are normally represented in conscious visual experience and the frequency with which human beings must manipulate or otherwise interact with these objects, it is natural to suppose that motor control systems should often avail themselves of consciously encoded visuo-spatial information in determining the spatial parameters of visually guided behavior. Indeed, Cutting and Vishton assume that the sources of depth information available for nearby targets are sufficient when suitably integrated to “produce the ability of human observers to accurately manipulate objects around them” (1995, p. 100). In what follows, we shall refer to the assumption that the spatial representational contents of visual experience are sometimes used to control actions directed at objects in the surrounding environment as the *Control Thesis*.

What notion of “control” is relevant to the Control Thesis? Morgan Wallhagen offers the following characterization:

...we can say that a mental state—particularly a sensory/perceptual state—plays a role in the control/guidance of some behavior when the state provides the information used by the motor system in developing motor instructions (e.g., information about joint movements) that bring about the behavior in question. Consider, for instance, reaching for one’s coffee cup. You glance over at your cup and reach for it, extending your arm, orienting your hand in a particular way, and configuring your fingers in a particular way. According to the present suggestion, your visual experience of the cup will have been utilized in the control and guidance of this behavior if it provides the information your motor system uses in causing your hand to move in the relevant direction and distance, to take on the appropriate orientation, and in configuring your fingers appropriately. (2007, p. 6)

This account is helpful, but it leaves to the side a distinction between two different *kinds* of visual information that the motor system can use in developing instructions for actions. For example, suppose that you are building a birdhouse, and you need to nail two pieces of wood together. To select the right tool for the job and to pick it up in the right way, you need categorical or “semantic” representations that help you to identify which object is the hammer, distinguish its handle from its head, and grasp it with a force appropriate to its apparent weight. But to grasp a target object in a way that conforms to its location, shape, and size, semantic representations of this sort are not really relevant. Instead, you need information about the *spatial layout* of the surrounding environment—in particular, bottom-up information about the target’s shape, size, direction, orientation, and distance in depth, computed on the basis of retinal input. Selecting and picking the hammer up successfully requires visual information of both kinds, so anyone whose capacities to acquire, store, access, or deploy either kind of information are impaired will

likely exhibit visuomotor deficits. However, the nature of their visuomotor deficits will vary precisely depending on the type of impairment that underlies them. If high-level object categorization capacities are impaired, then the subject's motor behavior will be abnormal with respect to "semantic" characteristics like which object is to be grasped or whether it is picked up in a manner appropriate to its function (Buxbaum, Schwartz, & Carew, 1997; Creem & Proffitt, 2001; Forde & Humphreys, 2000; Hodges, Bozeat, Ralph, Patterson, & Spatt, 2000). If, in contrast, early visuospatial processing capacities are impaired, then the behavior will be abnormal with respect to its spatial parameters (e.g., reaching location and grip scaling) instead. Although both kinds of information are typically used in controlling object-directed actions, it seems clear that their respective contributions to visuomotor processing are importantly distinct.

In light of this distinction, we can build on Wallhagen's characterization by distinguishing two roles that visual or visually based representations can play in behavioral control. First, bottom-up visual information about the spatial layout of the environment can be used for motor *programming*, that is, performing "the metrical computations that determine the precise trajectory and other spatiotemporal characteristics of the constituent movements" of an action (Milner & Goodale, 1995/2006, p. 244), both initially and if the action needs to be adjusted in flight (Goodale & Milner, 2004b). Second, high-level, categorical representations of the functional and material properties of objects that are not usable directly in motor programming can be used instead for action *planning*, that is, "choosing which target to act upon or deciding what type of action is most appropriate" (Milner & Goodale, 1995/2006, p. 244). A task for cognitive science is to explain the respective contributions of conscious and non-conscious visual representations to these two forms of behavioral control.¹

Articulated in these terms, the Control Thesis claims that conscious vision often plays a role in motor *programming*; that is, it claims that action-guiding systems draw on conscious visual information about the 3D spatial layout of the perceived scene, which information is then used in determining the spatial parameters of visually guided behavior, such as the direction in which you move your hand when reaching for an object or the size of the opening between your index finger and thumb. The thesis is non-committal, however, as to whether the contents of visual experience also supply the *categorical* representations utilized in motor planning. While it is relatively uncontroversial that visual experiences represent low-level visual attributes such as shape, size, distance, and direction, there is substantial disagreement about whether or not they also represent high-level, natural and artifactual kind properties such as *tiger*, *pine tree*, *hammer*, and *porcelain* (see Briscoe, 2015; Byrne, 2009; Jackendoff, 1987; Prinz, 2012; Siegel, 2010). The Control Thesis prescinds from this debate. It makes no claim about whether high-level categorical properties are sometimes represented in visual experience (as opposed to beliefs and judgments based on those experiences). Hence, it makes no claim about whether or not the contents of visual experience *per se* are capable of supplying visuomotor planning systems with information about those high-level properties.

The Control Thesis maintains that the spatial contents of visual experience are sometimes used in motor programming. Why do we say that visual experience plays this role

only sometimes rather than always? There are two related reasons. First, there is a large body of empirical evidence that non-conscious sources of visual information can significantly influence the spatial parameters of object-directed, bodily actions (see Jeannerod, 2006, chap. 3; Koch, 2004, chap. 12 for reviews). One series of important experiments, for example, involved shifting the location of a visible target during a saccadic eye movement (Bridgeman, Kirch, & Sperling, 1981; Goodale, Péliisson, & Prablanc, 1986; Pelisson, Prablanc, Goodale, & Jeannerod, 1986). Although subjects were completely unaware of the change in the target's location (they reported a single, stationary object), they automatically made correction saccades and rapid adjustments to pointing movements that accommodated its displacement. Further evidence that unconscious visual information can be used to control bodily actions adaptively comes from an experiment by Castiello, Paulignan, and Jeannerod (1991). Subjects were instructed to reach for one of three vertical dowels as soon as it became illuminated. In one condition, the middle dowel was illuminated, but when subjects began to reach for it, the light sometimes shifted unexpectedly to the dowel on the left or right, which then became the new target. Subjects were instructed to emit a vocal response as soon as they saw the central dowel illuminated and a second one if the position of the light changed. For both the initial hand movement toward the middle dowel and corrective hand movement toward the second dowel, action was found to precede visual awareness. Indeed, on some trials, "the dissociation between motor and vocal responses was such that subjects, to their own surprise, were already lifting the [second] dowel while they emitted the [second] vocal response" (Castiello et al., 1991, p. 2645). Experimental findings such as these have familiar, everyday counterparts: In many cases—think of the motorist who swerves to avoid the sudden appearance of an animal in the road or the tennis player who rapidly shifts her footing to receive to a fast serve—our actions begin to unfold before we consciously experience the objects or events to which they are responsive.

This leads to a second point, which is that there are reasons to think that the extent to which an action is consciously guided depends on several factors, including the type of action it is, the circumstances in which it is performed, and the kind of object in relation to which it is directed. When an intended action is simple or well-learned, performed in straightforward circumstances, and directed at a geometrically simple target, one is often able to execute the action successfully without needing consciously to monitor its course to keep it on track. That is, the Control Thesis is compatible with the idea that in such circumstances one is able to perform an action on "autopilot" (Pisella et al., 2000) or in "zombie" mode (Clark, 2007; Koch & Crick, 2001; Wu, 2013). In contrast, the Control Thesis holds that complicated, delicate, or unfamiliar actions, as well as actions that are performed in difficult circumstances or in relation to complex spatial layouts, require a special sort of oversight: To act successfully in these circumstances, one must attend more or less carefully to what one is doing and draw on one's conscious experience in controlling one's bodily movements. As Kent Bach writes:

[S]ome actions are more routine than others, requiring little if any attention, no decisions or flexibility during performance, hence minimal awareness. Indeed, if the action

is not only routine but repetitive, maximal efficiency may require minimal awareness—too much attention may slow one down and cause unwanted variations in performance. On the other hand, a novel, intricate action, like attempting an untried surgical technique, may require continuous, detailed, and accurate awareness of what one is doing. One must be alert to the slightest unforeseen contingency and be ready to alter one's course of action accordingly. (1978, p. 368)²

Similarly, Norman and Shallice (1986/2000) distinguish *willed* actions, which are “carried out under deliberate conscious control,” from *automatic* ones, which are performed “without paying active, directed attention” to them. They provide the following as examples of actions that are usually consciously controlled: (a) those that involve planning or decision making; (b) those that involve components of troubleshooting; (c) those that are ill-learned or involve novel action sequences; (d) those that are judged to be dangerous or technically difficult; (e) those whose execution will require overcoming a strong habitual response or resisting temptation (p. 377). Thus, while the Control Thesis postulates a role for conscious visual representations in visuomotor programming, the extent of this role is not invariant: The force of saying that conscious visuospatial information plays this role *sometimes* is to emphasize that it is especially to the extent that executing a behavior successfully requires top-down control over its spatial parameters, that motor programming makes use of consciously encoded visual representations.

Importantly, as Bach's talk of “minimal awareness” suggests, it would be a mistake to think that when performing very simple or routine actions a person acts *blindly*, without using *any* sensory information to guide his or her behavior. That this is not so is illustrated by a number of everyday examples, such as those of the skilled driver or tennis player, whose actions are at once automatic and clearly sensitive to visual perception of their environments. The relevant distinction is not between actions that are and are not informed by perceptual feedback, but between those actions that are guided only *implicitly* and *inattentively*, and those that are guided *deliberately* and by paying *attention* to what one is doing. And what the Control Thesis suggests is that especially in the latter case, though not necessarily only then, the spatial contents of conscious visual experience are used in programming the spatial parameters of our visually guided behaviors. It therefore predicts that impairments in conscious visual perception will lead to impairments in the performance of visually based actions that require, for any of the above reasons, more than “minimal awareness” for their successful performance.³

Spelled out in this way, the Control Thesis seems to capture a fairly intuitive, yet also empirically and theoretically motivated, conception of the relationship between visual experience and visually guided action. This conception, despite its intuitive appeal, is widely supposed to be contradicted by empirical evidence supporting the Two Visual Systems Hypothesis (TVSH) in visual neuroscience (Goodale & Milner, 1992, 2004a; Milner & Goodale, 1995/2006), according to which the spatial information used in visuomotor programming is supplied by a visual subsystem whose representations are entirely non-conscious. As A. David Milner and Melvyn Goodale write, according to their interpretation of TVSH “although we may be conscious of the actions we perform, *the visual*

information used to program and control those actions can never be experienced" (Milner & Goodale, 2008, p. 776; emphasis added)—a claim that contradicts the conception of visuomotor programming put forward by the Control Thesis. Our aim in this study is to show, contrary to this supposition, that the evidence put forward by proponents of TVSH supports an account of the role of conscious visual experience in bodily action that is entirely compatible with the Control Thesis. That is, we will argue that the available experimental evidence suggests that bottom-up spatial information present in *conscious* visual experience is sometimes used in motor *programming*, that is, determining the spatial parameters of the object-directed actions we perform.

Our study will proceed as follows. In Section 2, we motivate TVSH and summarize the main bodies of evidence for it, emphasizing a specific interpretation of the hypothesis that we will refer to as the *Perception-Action Model* (P-A Model). On this model, which we oppose to the Control Thesis, visuomotor programming draws exclusively on non-conscious visual information, and the behavioral role of conscious visual experience is limited to an influence on motor planning. In Section 3, we challenge the P-A Model on empirical grounds, proceeding in several stages: we argue first that evidence from patients with optic ataxia (Section 3.1) and profound visual form agnosia (Section 3.2) supports a significant role for conscious visual representations in motor programming, and that this hypothesis is further supported by studies of the behavioral effects of visual illusions (Section 3.3). In Section 3.4, we rebut two arguments that aim to show that conscious visual representations are unsuited for motor planning because they are exclusively allocentric. Section 4 offers some concluding remarks.

2. The two visual systems hypothesis

The two visual systems hypothesis is a reworking of Ungerleider and Mishkin's (1982) influential interpretation of the ventral processing stream from primary (V1) to inferotemporal cortex (IT) as a "what" stream devoted to object perception, and the dorsal processing stream from primary to posterior parietal cortex as a "where" stream devoted to spatial localization.⁴ On the original model, the theoretically fundamental distinction between the two processing streams is due to the difference in their respective contributions to the contents of conscious visual experience. The ventral stream is specialized for processing visual objects and their features, such as shape, pattern, texture, and color, whereas the dorsal stream is specialized for processing spatial relations between objects, including their locations relative to the perceiver. Importantly, on this model, the respective products of ventral and dorsal processing are seen as contributing to an integrated, general purpose conscious representation of the visible world, whose contents can be employed in thought and action alike.

In contrast to the Ungerleider–Mishkin model, TVSH postulates a quite different division of labor between the dorsal and ventral streams. According to TVSH, the crucial distinction is not between "what" and "where," but rather between "vision-for-perception" and "vision-for-action." On this account, the neural correlates of the contents of conscious

visual awareness reside exclusively in the ventral processing stream, whose functional role is limited to categorizing (recognizing) objects and events, determining their task-relative significance, and keeping track of their locations in a scene-based or allocentric frame of reference. Content-specific differences in visual awareness, for example, whether an object is represented as red or green, near or close, moving or stationary, supervene entirely on differences in neural activity in this stream. Furthermore, the behavioral utility of these conscious representations is restricted to a role in *motor planning*: Spatial information in the ventral stream is not used in programming the spatial parameters of bodily actions. Instead, it is the dorsal processing stream that supplies the spatial information involved in real-time visuomotor programming and, accordingly, represents object locations in viewer-relative or “egocentric” frames of reference. TVSH holds that this stream mediates allocation of attention to consciously perceived objects and their features but otherwise functions unconsciously as a “zombie” system (Clark, 2007; Koch & Crick, 2001).

For our purposes, the crucial claim at issue here is that the bottom-up visual spatial information used by motor systems in programming the spatial parameters of behavior is drawn *exclusively* from the non-conscious dorsal stream, whereas bottom-up spatial information originating in the ventral stream never plays this role. The role of the ventral stream—and so conscious vision—is limited to motor planning, that is, supplying high-level, semantic information about the material, functional, and task-relative properties of target objects, and not the egocentric spatial information used in setting the spatial parameters of selected actions. Thus, Goodale and Milner draw the conclusion that the “intuitively obvious” idea that “the visual image that allows us to recognize a coffee cup is the same one that guides our hand when we pick it up” is based on an “illusion,” as “the visual system that gives us our visual experience of the world is *not* the same system that guides our movements” (2004a, p. 2). On this account, what conscious visual experience provides is just a representation that permits us to recognize the cup *as* a cup, and thus to direct our actions toward it in task-appropriate ways, for example, reaching for it when we need a drink, picking it up by the handle instead of the rim, and grasping it lightly if it appears to be hot. The spatial parameters of these movements, however, are programmed in a way that draws entirely on non-conscious visual information.

There are two main sources of empirical evidence for this account, which we will summarize here before examining them in more detail below. First, Milner and Goodale argue that their hypothesis is best able to account for the pattern of deficits and spared abilities in patients with lesions affecting one or the other of the two streams. Patient DF, with extensive bilateral damage to the lateral occipital area (LO) of the ventral stream, can visually experience the colors, textures, locations, movements, and stereoscopically defined depths of perceived objects but is severely impaired in consciously perceiving their shapes and orientations (Milner & Goodale, 1995/2006; Milner et al., 1991). Nonetheless, DF is able to perform many visuomotor tasks with surprising fluency. When asked, for example, to grasp an object or to post a card through a slot, she is almost indistinguishable from normal controls. In contrast, subjects with optic ataxia, caused by damage to superior parietal areas in the dorsal stream, have normal visual acuity in foveal

areas and normal recognitional abilities, but experience difficulties in the visual control of hand and arm movements directed at objects, especially those in the periphery of the contralesional visual field (Perenin & Vighetto, 1988). For example, optic ataxics may miss target objects in vision by a few centimeters when reaching in their direction, and they show poor scaling of grip and hand orientation when attempting to take hold of them. Furthermore, in reaching tasks, optic ataxics have difficulty in making fast corrective movements to a jump in target location (Battaglia-Mayer & Caminiti, 2002).⁵ TVSH can account for this differing pattern of deficits and spared abilities between subjects with ventral and dorsal damage, as it maintains that the spatial information used to support world-engaging action is computed independently of the information used to support conscious object perception and identification.

The second primary source of evidence for TVSH comes from behavioral studies of the comparative effects of visual illusions on visual awareness and visuomotor action. A now-famous study by Aglioti, DeSouza, and Goodale (1995) involves a graspable version of the famous Ebbinghaus (or Titchener Circles) illusion, in which two central circles are presented, each surrounded by an annulus of other circles (see Fig. 1). In one case, the surrounding circles are larger than the central one, whereas in the other, they are smaller. Aglioti et al. (1995) constructed a 3D version of the illusion, using thin solid disks. Subjects were asked to pick up the central disk on the left if the two central disks appeared identical in size and to pick up the central disk on the right if they appeared different in size. The experimenters varied the relative sizes of the two target disks randomly so that in some trials physically different disks appeared perceptually identical in size, whereas in other trials physically identical disks appeared perceptually different in size. In selecting a disk in either trial condition, Milner and Goodale observe, “subjects indicated their susceptibility to the visual illusion” (1995/2006; p. 168); that is, their choice of which disk to pick up was determined by its apparent size rather than its real one. Nonetheless, the effect of the illusion was significantly less pronounced with respect to action, as measured by maximum grip aperture (MGA) in prehension, than with respect to conscious perception, as measured by the distance between thumb and forefinger in a manual estimate of disk size. Similar findings have been reported for a variety of other visual illusions, including the Müller-Lyer illusion (Daprati & Gentilucci, 1997), the Ponzo illusion (Jackson & Shaw, 2000), the Dot-in-Frame illusion (Bridgeman, Peery, & Anand, 1997), and, recently, the Hollow-Face illusion (Króliczak, Heard, Goodale, & Gregory, 2006). In each case, there is a measurable divergence between what subjects consciously see and the way their behavior unfolds: the spatial parameters of visually guided action seem not to be determined by the illusory spatial content of conscious visual experience.

Importantly, all this evidence exists alongside findings that support a role for the ventral stream in action *planning*; that is, identifying or choosing among possible actions (e.g., selecting which disk to reach for in the 3D Ebbinghaus task), planning complicated action sequences that reflect the subject’s intentions or goals, and providing “semantic” information that influences how actions are performed (as in our example of the temperature of the coffee cup). Thus, for example, when instructed to grasp an object (like a screwdriver, say) whose functional properties determine how it should be picked up (here,

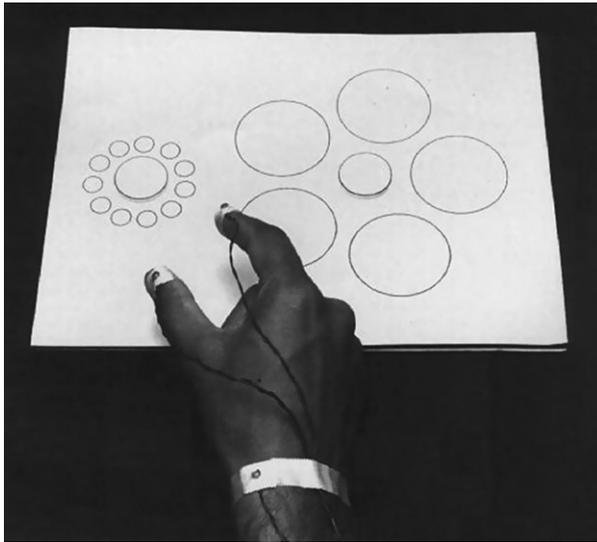


Fig. 1. 3D version of the Ebbinghaus illusion. Aglioti et al. (1995) found that the way subjects scaled their grip in reaching for a target object was affected less by the illusion than was their conscious perception of object size.

by its handle), DF picks the object up in the right way only if she is able to infer its identity based on surface cues, such as the textures and colors of which she remains consciously aware (Carey, Harvey, & Milner, 1996). Similarly, it has been found that, while DF scales her grip appropriately to the size of a target object, the force of her grip will not be determined by the object's weight unless, again, this can be inferred from its consciously visible surface features (McIntosh, 2000). Again, however, the finding that information present in the ventral stream can significantly modulate action planning with high-level information about an object is consistent with the denial that the spatial contents of visual awareness are ever used in programming the trajectory of motor behavior: instead, their role is supposed to be limited to the forms of action planning described above.

On the basis of these and other such findings, proponents of what we will term the “P-A Model” of visuomotor control conclude that the role of conscious visual experience in guiding motor action is restricted to the selection of targets and the provision of high-level categorical information that affects certain aspects of visually guided behavior. The P-A Model postulates a strict division of labor between the dorsal and ventral streams, analogous to that between a remotely assisted robot (Goodale & Humphrey, 1998, pp. 202–203; Goodale & Milner, 2004a, pp. 98–101; Milner & Goodale, 1995/2006, pp. 231–234) or heat-seeking missile (Campbell, 2002, pp. 55–56) and its operator. On this model, the role of visual consciousness in the control of action is limited to selecting targets to engage on the basis of their task-relevant significance, and identifying types of actions to perform.⁶ Once these initial parameters are set, motor programming systems use their

own sources of non-conscious visual information about the environment's spatial layout to achieve the designated goal. Thus, Milner and Goodale write, in two passages that are worth quoting at length:

The role of the ventral stream in action . . . is to provide visual information to enable the identification of a goal object such as a coffee cup, and to enable other cognitive systems to plan the action of picking up that cup. This would include the selection of the class of hand postures appropriate to the particular task at hand (whether that be taking a sip of coffee, for example, or putting the cup in the sink). But action planning of this sort is quite abstract, and the final movements that constitute the action could take many different forms. It is the dorsal stream's job to use the current visual information about the size, shape, and disposition of the object in egocentric coordinates (in the case of the coffee cup, with respect to the hand) to program and control the skilled movements needed to carry out the action. This then is the specialized meaning we give to "vision-for-action": not the use of visual information for abstract planning, but rather its use in the detailed programming and real-time control at the level of elementary movements.

. . . The visual information used by the dorsal stream for programming and on-line control, according to the model, is not *perceptual* in nature . . . [I]t cannot be accessed consciously, even in principle. In other words, although we may be conscious of the actions we perform, the visual information used to program and control those actions can never be experienced . . . (Milner & Goodale, 2008, pp. 775–776)

And again:

. . .when we reach out to pick up an object, we may have full visual awareness of our arm moving and our hand configuring in a certain way, of the target object as having a certain shape and lying in a certain location, of the presence and nature of neighboring objects and so on. In other words, we may be fully visually aware of our actions and of the detailed environmental context in which they are made. But the essence of Milner & Goodale's (1995/2006) interpretation is that this visual awareness accrues not from visuomotor processing in the dorsal stream, but from concurrent processing in the ventral stream. Conversely, according to the model, *such ventral stream processing plays no causal role in the real-time visual guidance of the action*, despite our strong intuitive inclination to believe otherwise (what Clark 2001 calls "the assumption of experienced-based control"). (Milner, 2012, p. 2296, emphasis added)

If what Milner and Goodale say in these passages is correct, then the Control Thesis clearly is not. For according to that thesis, the role of conscious visual experience in visuomotor control goes beyond "flagging the target and selecting an action type, leaving the dorsal servants to do the rest" (Clark, 2007, p. 576): The Control Thesis says that conscious visual representations are also used in the metrical computations that generate

motor instructions and guide our behavior as it unfolds. According to the P-A Model, the role of conscious representations in visuomotor guidance falls short of the sort of visuomotor guidance that we identified as central to the Control Thesis, as the model holds that the spatial contents of visual awareness are *not* utilized in programming the spatial parameters of visually guided behavior.

In what follows, we will take for granted that there are non-conscious systems in the human dorsal stream specialized for visuomotor guidance, and that most of the neural correlates of visual consciousness reside in the ventral stream.⁷ Our question will be whether visually based motor programming always draws *solely* on information from these non-conscious systems, as supposed by the P-A Model, or whether instead, as the Control Thesis has it, the spatial information present in conscious vision also has a significant role to play in motor programming, especially in cases of difficult, complex, or non-routine behaviors. We review experimental evidence supporting the latter position, that is, that the Control Thesis is compatible with a proper understanding of TVSH after all. In contrast, the radical perception-action dichotomy of the P-A Model will turn out to be a poor fit for the evidence we discuss.

3. A skeptical assessment

Section 2 presented evidence for an understanding of the human visual brain as composed of two subsystems, one responsible for generating our conscious experience of the world and the other specialized for motor programming, including the on-line guidance of action. That evidence breaks down into two main categories:

- *A dual dissociation:* Despite significant ventral stream damage and corresponding deficits in conscious visual perception, patient DF is capable of many visually guided actions. In contrast, despite relatively normal conscious visual percepts, patients with lesions of the posterior parietal cortex often experience difficulties in visually guided arm and hand movements (optic ataxia).
- *Illusion studies:* Certain visually guided behaviors seem to be relatively independent of visual illusions; for example, the way one scales one's hand to grasp a disk in a 3D version of the Ebbinghaus illusion is closer to its real size than its illusory one.

We will argue in this section that, when closely examined, both sources of evidence are compatible with the Control Thesis but in conflict with the starkly dichotomous P-A Model.

3.1. *Optic ataxia*

We turn first to evidence from studies of patients with optic ataxia. As Milner and Goodale argue, reaching errors in subjects with optic ataxia show that consciously encoded visual information concerning the shape, orientation, and location of target

objects is not by itself *sufficient* for normal visually guided action. By itself, however, this does not conflict with the Control Thesis, which maintains only that conscious visual experience is *a* source of the spatial information used in visuomotor programming. As such, the Control Thesis is compatible with the observation that adept visually guided action requires that conscious visuospatial information in the ventral stream (like non-conscious visuospatial information in the dorsal stream) be subjected to sensorimotor transformation. For example, visual representations of an object's location relative to the perceiver's *eye* must be transformed into a representation of the object's location relative to her *hand*, if the former are to guide reaching and grasping. To perform this transformation, eye-relative representations must be calibrated with up-to-date proprioceptive information about the spatial configuration of the perceiver's body, in particular, the current location of her arm and hand (for an overview, see Shadmehr & Wise, 2005).

Moreover, experimental evidence suggests that optic ataxia may be caused in part by a proprioceptive deficit with respect to the location of the contralesional hand, independent of any deficit in the visual perception of the target. Blangero et al. (2007) found that patients with optic ataxia made significant errors when pointing in the dark with their ipsilesional, normal hand to the location of their contralesional, ataxic hand, and vice versa (also see Granek, Pisella, Blangero, Rossetti, & Sergio, 2012). It is thus possible that optic ataxia involves a breakdown in systems that monitor hand position and/or integrate proprioceptive hand position signals with visual target information. If so, then given the need for accurate proprioceptive information in sensorimotor transformation, the finding that patients who suffer from optic ataxia have difficulty in controlling reaching and grasping movements using their contralesional hand would reveal nothing at all about the role of conscious vision in programming grasping actions.

In short, the Control Thesis does not entail that consciously encoded spatial information should be by itself sufficient for adept visuomotor action, as the role it postulates for visual experience in visuomotor guidance is fully consistent with the idea that to guide action, information about an object's spatial attributes present in visual awareness must be integrated "downstream" with information about the configuration and posture of the subject's body. Hence, the Control Thesis is not at all contradicted by evidence that lesions to parietal areas will impair visuomotor performance, even as conscious vision is spared.

Before moving on, a further point is germane. Optic ataxia typically impairs visual control of hand and arm movements directed at objects in the periphery of the contralesional visual field (Perenin & Vighetto, 1988). "Pure" optic ataxics with unilateral lesions are often able accurately to perform hand and arm movements targeted on stationary objects in central vision. In a recent study, Hesse, Ball, & Schenk (2012) found that DF's pointing and grasping behavior is significantly impaired for targets in peripheral vision. In other words, DF's visuomotor deficits are quite similar to those observed in patients with optic ataxia. This seems to conflict, as Hesse and colleagues emphasize, with the claim that the pattern of deficits and spared abilities, respectively, characteristic of visual form agnosia and optic ataxia form a double-dissociation.⁸

3.2. *Profound visual form agnosia*

As we have seen, patient DF is put forward as an example of someone with intact capacities for visually guided action despite impairments in conscious vision, and hence as evidence that conscious vision does not play a significant role in motor programming (as opposed to motor planning). Milner and Goodale have always emphasized that the two cortical streams must ultimately work together in the production of purposive, world-engaging actions, and that substantial interaction between the two streams is functionally necessary for a wide variety of familiar actions involving tools and other “meaningful” artifacts. As they put it, “the dorsal system by itself can ensure that [an object] is grasped in a metrically accurate manner, but only the ventral stream can ensure that it is picked up at the right end” (Milner & Goodale, 1995/2006, p. 229). On this view, conscious visual representations in the ventral stream play an important role in selecting targets for visuomotor engagement and types of action to perform on the basis of categorical or “semantic” information retrieved from long-term memory. It is the dorsal stream’s job alone, however, to determine the spatial parameters of selected actions on the basis of its own sources of bottom-up information about the spatial properties of target objects. And this is supposed to be demonstrated by the case of DF: Despite her conscious visual impairments, she is able to perform actions with surprising fluency on the basis of unconscious spatial information originating in the dorsal stream.

However, just as we observed in Section 3.1 that the Control Thesis does not hold that conscious visual information is on its own *sufficient* for normal visually guided action, so it is important to recall that it does not entail that conscious vision is always *necessary* for skillful visuomotor action, either. For the Control Thesis does not claim that it is the proprietary functional role of conscious vision to guide our object-directed actions. As we emphasized above, the thesis takes for granted that while some actions require conscious oversight or monitoring for their successful performance, for example, because they are complex or unfamiliar or otherwise difficult to perform, other actions can be performed without such oversight in “autopilot” mode. So the possibility of profound visual form agnosia does not by itself present any general, empirical challenge to the Control Thesis. The fact that DF has non-trivial residual capacities for visually guided action is compatible with the Control Thesis, *so long as those capacities only extend to actions that can be performed by normally sighted subjects in a routine or automated fashion*. The crucial questions are thus the following:

1. Is DF adept in performing visually guided actions even in those contexts where the Control Thesis predicts that conscious oversight will be necessary?
2. If not, then are her behavioral deficits in those contexts due to failures in motor *planning*, and thus due to DF’s inability to access high-level semantic or categorical visual representations, or are they also due to failures in motor *programming*?

In the remainder of this section, our strategy will be to show that the answer to the first question is unambiguously negative, and, further, that, in answer to the second question, many of the observed deficits in DF’s visuomotor performance are not plausibly

interpreted as being due just to impaired motor planning. This will support the claim, central to the Control Thesis, that conscious visuospatial representations are sometimes drawn on in programming the spatial parameters of object-directed actions.

To begin, we should note that numerous studies have called into question the idea that DF's visuomotor performance is always fully normal even in very simple tasks. Though she is clearly far *more* proficient at visually based "motor" tasks like reaching, grasping, and posting than "perceptual" ones like making visually based judgments or manual estimations of orientation, location, or size, DF's performance in visuomotor tasks differs from that of healthy control subjects in conditions where perceptual input is compromised or impoverished; for example, under monocular viewing conditions (Carey, Dijkerman, & Milner, 1998; Dijkerman, Milner, & Carey, 1998, 1999; Marotta, Behrmann, & Goodale, 1997), when visual depth information is distorted by prisms (Mon-Williams, McIntosh, & Milner, 2001), and when target objects are presented in peripheral vision (Hesse, Ball, & Schenk, 2012). All this suggests that even in those conditions where DF's visuomotor behavior *appears* fully normal, the strategies she uses to program her actions may be different from those of normal subjects with a functioning ventral stream. As Schenk (2010) observes, this suggestion is further supported by the fact that, in contrast with DF, healthy subjects are able to act relatively normally despite the removal or distortion of binocular depth cues (Marotta et al., 1997; Mon-Williams et al., 2001; Wann, Mon-Williams, McIntosh, Smyth, & Milner, 2001), revealing that there is some redundancy in the visual information ordinarily used for motor programming. Given this, even if DF were completely normal in her visuomotor performance despite her ventral stream lesions, it would not follow that healthy subjects do not use information from the ventral stream in motor programming, for DF might simply be getting by without it.

As we indicated above, however, the Control Thesis emphasizes the role of conscious visual representations in motor tasks that are targeted on geometrically complex stimuli or that are unfamiliar or otherwise difficult to perform, as it hypothesizes that it is in tasks like these that conscious visual information will be used most extensively in motor programming, due to the need for attentive oversight. Here again, there is evidence from patient DF that supports our prediction that the programming of complex actions will be compromised when conscious vision is impaired. DF's performance has been found to be far from normal when she is required to engage objects with geometrically complex shapes, as she has trouble determining the contact points for grasping an X-shaped object or attempting to post a T-shaped object through a T-shaped slot (Carey et al., 1996; Goodale et al., 1994). In another study, worth discussing in detail, DF was instructed to reach out and pick up circular disks by placing her fingers through holes cut in them (Dijkerman et al., 1998). Each disk had either two or three holes, one of them larger (22.5 mm diameter) and intended for the thumb, another one smaller (20 mm) and intended for the forefinger, and in the "three-hole" condition a third hole also 20 mm in diameter and intended for the middle finger (see Fig. 2). To make the task challenging, within each condition the disk orientations were varied, as were the locations of the holes, creating numerous configurations to which subjects needed to be sensitive when they reached for each stimulus. In each case, subjects were instructed to open their eyes, and then "reach

out to grasp the disk by placing the three [or two] fingers into the holes as quickly or accurately as possible” (Dijkerman et al., 1998, p. 428), without any specific instructions given as to which finger was supposed to go into which hole—however, the thumb hole was always located toward the bottom of the stimulus, indicating a “natural” mode of grasping in each case.

Fig. 3 below illustrates DF’s performance in the three-hole task, both when all three holes were the same color (open circular symbols) and when they were painted three different colors (open diamond symbols). Several findings should be emphasized:

- As the left-hand graph illustrates, in the “same-colored” condition the size of DF’s grip was entirely insensitive to the distance between the thumb and forefinger holes: There was no statistically significant difference in her grip aperture between

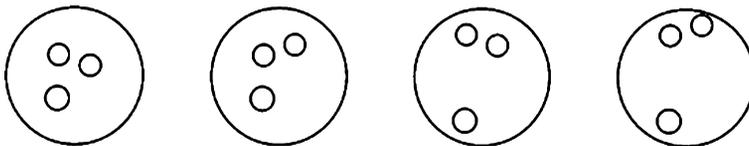


Fig. 2. Sample stimuli from the “three-hole” conditions in Dijkerman et al. (1998). The distance between the thumb (bottom) and forefinger holes and relation of the middle finger hole to the forefinger hole were varied, and the disks could be oriented vertically or tilted 45° to the left or right. Subjects were instructed to reach out and grab the disk by its holes “as quickly or accurately as possible” (p. 428).

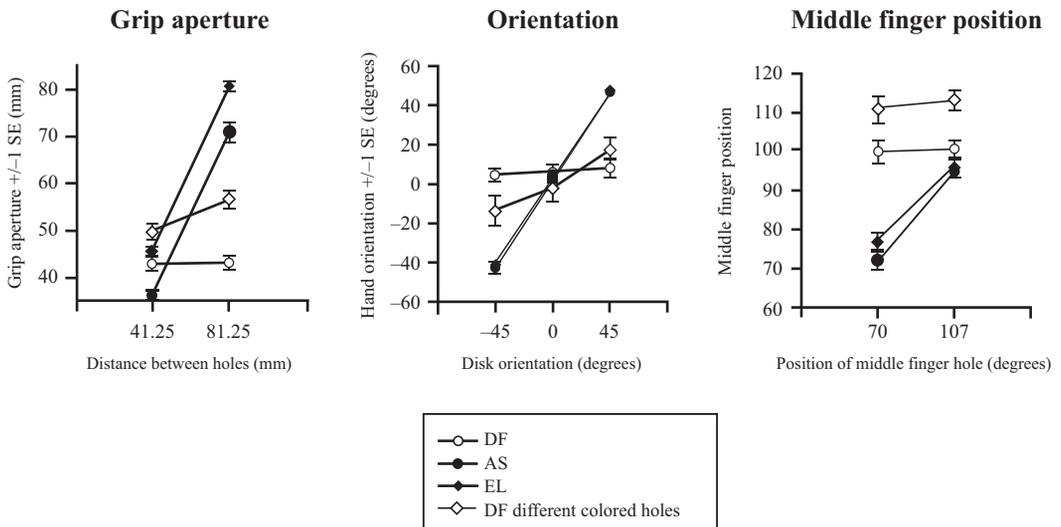


Fig. 3. Performance of DF (open symbols) and two normal subjects (colored symbols) in the “three-hole” conditions in Dijkerman et al. (1998). The left-hand grasp shows the separation between thumb and forefinger, the middle grasp the orientation of the hand, and the right-hand grasp the position of the middle finger relative to the forefinger and thumb. All data are from 40 ms before contact with the disk.

conditions where the holes were 41.25 and 81.25 mm apart. This improved somewhat in the “different-colored” condition, where DF’s grip aperture varied significantly between the two hole distances, but it was still quite impaired in comparison to normal subjects.

- Similarly, as the center graph illustrates, in the three-hole same-colored condition there was no significant difference in DF’s hand orientation depending on disk orientation: The configuration of the stimulus (either vertical or 45° to the right or left) did not affect the way she turned her wrist and hand to pick it up. Again, this improved somewhat in the different-colored condition, but her performance was still “far from normal” (Dijkerman et al., 1998, p. 432).
- Moreover, as the right-hand graph illustrates, in both conditions involving the three-holed disks there was no relationship at all between the position of DF’s middle finger and the position of the hole intended for it (which, again, could be either 10 mm above or below the forefinger hole).
- Finally, on 55.6% of the same-colored trials and 33.8% of different-colored trials DF inserted the inappropriate finger into a given hole (e.g., her forefinger into the hole intended for the thumb), a mistake that control subjects never made.

All these aspects of DF’s performance were markedly inferior to control subjects, as well as her performance in the simpler grasping tasks discussed above.

As Fig. 4 illustrates, DF’s performance was much better in the two-hole conditions, where she adjusted her hand to the disk orientation and location of the holes (left, right, or parallel to the diametric axis) almost as accurately as control subjects. However, DF still placed her fingers in the wrong holes 39.5% of the time, and her grip aperture was only barely sensitive to the distance between the holes. Coloring the holes differently improved these aspects of her performance only slightly: There was a highly significant effect of interhole distance on grip aperture, but still no significant difference between the

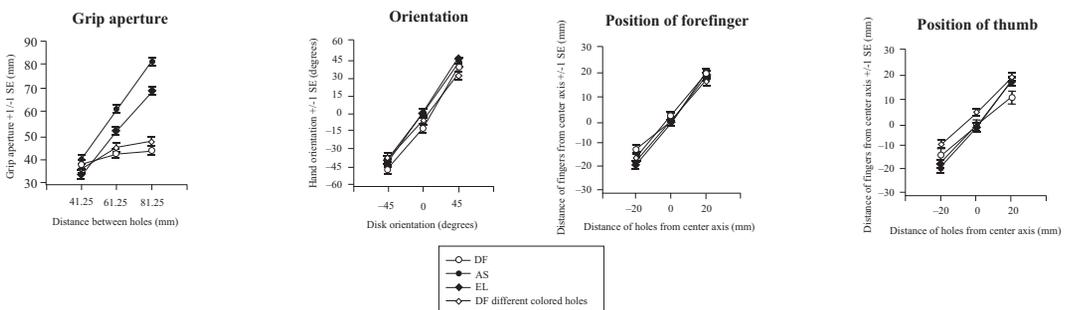


Fig. 4. Performance of DF (open symbols) and two normal subjects (colored symbols) in the “three-hole” conditions in Dijkerman et al. (1998). The leftmost grasp shows the separation between thumb and forefinger, and the others show the orientation of the hand and the accuracy in aiming for the part of the disk (left, right, or center) where the holes are located. In these latter three respects DF’s performance was almost as good as that of the control subjects.

intermediate (61.25 mm) and large (81.55 mm) distances. She also inserted her fingers into the wrong holes less frequently, but still 9.9% of the time.

To sum up, then: in sharp contrast with her ability to scale and orient her grip to the size of novel objects when reaching to grasp them by their periphery, DF's grasping behavior showed much less of this sensitivity in the present task, which required grasping objects through holes cut in them. That is, she was consistently quite insensitive to changes in hole location and disk orientation when grasping disks with three holes, and to changes in interhole spacing when grasping disks with two holes. Moreover, in each condition she frequently inserted her fingers into the wrong holes—a mistake that normal subjects never made.

How should we account for DF's poor performance on these tasks, in comparison to the other visuomotor tasks in which she has been found to be so proficient? The authors of the study interpret their findings as providing support for the claim that DF's intact dorsal stream represented objects in exclusively egocentric coordinates, and she lacked access to the allocentric or scene-based representations that were required for success in the task. They claim that their task “was designed to force the use of allocentric spatial processing, while at the same time making it impractical to use egocentric spatial information” (Dijkerman et al., 1998, p. 425). On their account, the purely egocentric spatial information present in DF's intact dorsal stream “would have sufficed to allow her to direct her hand to the appropriate left/right part of the disk and, arguably, also to orient her hand with respect to a single pair of holes” (ibid., p. 433). However, this information did not suffice to determine the appropriate scale of her grip, and so “this ability to orient the hand was lost” when a third hole was present (ibid.). As they write:

... successful execution of the complete act of disk-grasping requires an allocentric coding of the pattern of holes, and thereby the participation of the ventral as well as the dorsal stream. The dorsal stream on its own can guide the forefinger (usually the leading digit) accurately into one of the holes and can also cause the wrist to rotate to the correct orientation formed by the two holes. This could all be readily achieved using egocentric coding. But the opening of the forefinger-thumb grasp, as well as the choice of the correct hole for the forefinger to enter, would most effectively be guided by coding the *relative* hole positions on the disk. If she did not have the benefit of such allocentric coding, DF would be severely hampered in performing those elements of the task. (Dijkerman et al., 1998, pp. 433–434)

We believe this account of DF's performance in terms of a deficit in allocentric coding is unsupported. It seems plausible as an explanation of her tendency to insert her fingers into the wrong holes in the two- and three-hole tasks, as it may be that it is only by allocentrically representing the three holes' relative sizes that the subject in this task can determine *which* finger belongs in each hole, similar to the finding by Dijkerman et al. (2009) that DF was deficient in planning an appropriate grasp posture for a prism-shaped object presented at different orientations. However, it is very unclear what independent reason there is to think that anything similar can explain DF's inaccurate hand rotations

and grip sizes, as programming the proper grip size and hand rotation for grasping an object appears to be a paradigmatically egocentric achievement, requiring only a representation of the target's position in a body-centered space: In the present task, all that matters is the location of each hole relative to the finger that has been designated to go in it.⁹ If grasping an unfamiliar object by its periphery only requires representing the location and shape of the object with respect to the position of one's hand and fingers, it seems that grasping a disk through holes cut in its surface should be achievable by similar means: Spacing one's fingers by the appropriate distance should not require representing *relative* hole positions in this task any more than grasping something by its periphery requires representing the relative orientations of its sides, as opposed to their positions in a hand-centered space. There is no evident reason to think that egocentric representations are not enough to determine how to rotate one's hand, or how far apart one's fingers need to be, to insert them into two or three circular holes cut in a disk, nor is there any clear reason why they would be more essential to the two-hole task than the three-hole one. Aside from its potential to bring these data into line with the P-A Model, the interpretation of these impairments as simply the product of a deficit in allocentric coding is unconvincing.¹⁰

In contrast to the P-A Model, the Control Thesis offers a natural explanation of these findings. This is that unlike grasping a simple object by its periphery, the task of picking up an object through holes cut in it is relatively *complex and unfamiliar*, and therefore demands a particular sort of conscious oversight to be executed accurately. This is acknowledged by the authors of the study, who write that “even healthy subjects do not find the task [of grasping an object through three holes] a normal or natural one to perform—it does not have the effortless feel of other everyday visuomotor tasks, such as grasping solid objects. Indeed, introspection suggests that *grasping the disks used in this study requires conscious monitoring of the whole prehension process in a rather deliberate fashion*” (Dijkerman et al., 1998; p. 434; emphasis added). According to the Control Thesis, this conscious monitoring is necessary because of the complexity and unfamiliarity of the task, and thus it predicts that performance in such a task will be impeded when conscious vision is impaired, as it is in patient DF.¹¹ Thus, we submit that DF's performance in this experiment further supports the conception of visuomotor action captured in the Control Thesis.

In sum, the findings reviewed in this section support an affirmative answer to our question (1) above: DF's visually guided behaviors are not fully normal, especially in those contexts where the Control Thesis predicts that conscious oversight will be necessary. Taken in itself, this would pose no problem for the P-A Model if it could be shown in response to our question (2) that her visuomotor impairments are due only to a lack of categorical or allocentric visual representations, but we have argued that such an interpretation of these findings is ungrounded. Though she is clearly deficient in motor planning as well, some of DF's visuomotor impairments seem also to be grounded in poor programming of the egocentric spatial parameters of her behaviors, suggesting a role for conscious visual information in this process—a finding that is in line with the Control Thesis, but contrary to the predictions of the P-A Model.

3.2.1. *The function of ventral area LO*

As we noted above, DF's inability to perceive shape and orientation results from bilateral damage to the lateral occipital area (LO) of the ventral stream. fMRI findings suggest that computations in this area are poised between intermediate-level feature processing and high-level, shape-based object recognition (Kanwisher, 2004). In particular, while activations in LO vary with differences in two- and three-dimensional shape and are highly invariant with changes in target size and position, they do not display viewpoint invariance and do not appear to be selective for specific categories of objects or individuals (Grill-Spector et al., 1999).

We propose that the evidence summarized in the preceding section that DF's spared visuomotor competence is limited to programming rudimentary actions targeted on geometrically simple objects, when considered together with her perceptual deficits, suggests that spatial information originating in area LO contributes both to conscious visual perception and to the control of real-time visuomotor action in normal subjects. In particular, it suggests that LO contributes viewer-relative information about a target's shape that augments the dorsal stream's own bottom-up sources of input. While it is clear that not *all* grasping is dependent on contributions from LO—as indicated, for example, by DF's ability to pick up small disks and rectangles accurately—her inability to perform visuomotor actions involving objects with more complex geometrical shapes suggests that contributions from LO to such actions are far from negligible.

In contrast to this interpretation, Culham et al. (2003) provide evidence that might seem to challenge a direct role for area LO in visuomotor programming. In an fMRI study, they compared neural activations in dorsal area AIP and the lateral occipital complex (LOC)—which contains area LO—during grasping and reaching trials. Culham et al. (2003) found that dorsal area AIP is strongly activated during grasping, when object size and shape computations are required to preshape the hand, but is much less activated during reaching, which does not require preshaping. Activations in LOC, in contrast, were found to be stronger during both grasping and reaching trials than during passive viewing, but not to *differ* between them. These results, Culham and colleagues suggest, support the view that the computations of object properties for hand preshaping do not require any involvement of the ventral stream: “it seems that LOC is neither [differentially] activated by, nor necessary for, object grasping” (Culham et al., 2003, p. 187).

But this conclusion is hasty. That areas in LOC do not show differential activation between reaching and grasping trials indicates only that they are not involved in *transforming* visual shape representations into object-directed motor actions. The absence of differential activation, however, does not conflict with the supposition that areas in LOC (such as LO) *supply* shape information to areas in the dorsal stream (such as AIP) that are involved in such visuomotor transformation. In this case, activation in LOC would remain the same, whereas areas downstream that receive information from LOC would exhibit different responses as a function of the specific motor tasks being undertaken. The best way to show that LO is not involved in supplying information used in motor programming would be to marshal neuropsychological evidence that actions targeted on

complex, geometrical stimuli are not impaired by lesioning or inactivating LO. But as we have argued at length, the available evidence from patient DF suggests that precisely the opposite is the case.

3.3. *Illusion studies*

The illusion studies discussed above indicate that visuomotor action is sometimes less susceptible to illusion than visual awareness. Proponents of the P-A Model draw from this the conclusion that spatial information present in conscious visual awareness is not the same spatial information used to program visuomotor actions. Andy Clark, for example, writes of a “deep and abiding dissociation between the contents of conscious seeing, on the one hand, and the resources used for the on-line guidance of visuomotor action, on the other” (Clark 2001, p. 495).

Circumspect examination of the empirical evidence, however, suggests a dissociation that is neither deep nor abiding. Rather, it suggests a far more qualified picture on which action *is* less susceptible to illusion than visual awareness, but only under specific conditions—notably including those predicted by the Control Thesis. In particular, action must be:

- *Rapid*: Actions that are performed slowly are not recalcitrant to visual illusions (Carey, 2001; Rossetti et al., 2005). Króliczak et al. (2006) found that even the high-level Hollow-Face illusion, in which a concave mask appears to be convex when illuminated from below, has a strong effect on slow pointing movements targeted on magnets affixed on the facing surface of the mask. “This suggests,” they write, “that pointing and other more deliberate and slow movements do not have to engage the ‘automatic’ visuomotor mechanisms in the dorsal stream but instead can be mediated by ‘perceptual’ processing in the ventral stream” (Króliczak et al., 2006, p. 15). Whether or not visuomotor programming in the dorsal stream is able to integrate spatial information present in conscious vision thus depends, in part, on temporal constraints on task performance.
- *Easy and well-practiced*: Gonzalez, Ganel, Whitwell, Morrissey, and Goodale (2008) found that although a display containing the Ponzo illusion does not have an effect on the scaling of a normal precision grasp, it does have a significant effect on the scaling of the awkward and unfamiliar grasping movements. “Grip scaling of the awkward movements, just like perception,” they write, “is fooled by the illusory display” (Gonzalez et al., 2008, p. 629). Their findings provide support for the view that “the more practiced or ‘automatic’ an action, the more likely that action is to engage the encapsulated visuomotor mechanisms in the dorsal stream . . . Conversely, the more deliberate and less practiced the action, the more likely it is that cognitive mechanisms, which do make use of ventral stream processing, would be involved” (ibid., p. 625).
- *Performed under binocular viewing conditions*: In the absence of binocular depth information, for example, stereopsis and convergence, target-directed grasping is

fully affected by geometrical visual illusions (Marotta, DeSouza, Haffenden, & Goodale, 1998; Goodale & Milner, 2004a, p. 92). Visuomotor actions undertaken by subjects who are blind in one eye or suffer from strabismus are presumed to rely on spatial information present in conscious vision.

- *Performed with the right-hand:* Gonzalez, Ganel, and Goodale (2006) instructed right- and left-handed subjects to reach and grasp small objects embedded in the Ponzo and Ebbinghaus illusions with either their dominant or their non-dominant hand. For both right- and left-handed subjects, they found that precision grasping with the left hand, but not with the right, was influenced by the illusions. “Our demonstration that the right hand shows greater resistance to size-contrast illusions than the left,” they write, “provides compelling evidence that the visuomotor networks controlling grasping (in the right hand) make use of visual information that is different from that used by the networks driving our [conscious] perception” (Gonzalez et al., 2006, p. 3500). Yet the evidence they adduce also suggests that the visuomotor networks controlling grasping with the left hand make use of the same sources of spatial information as are used by the networks driving conscious perception.

Outside of these conditions, visuomotor programming appears to be fully susceptible to the effects of visual illusions, suggesting that it is dependent on sources of bottom-up spatial information present in conscious vision (for further discussion, see Briscoe, 2008a, 2009). While this assessment speaks against a sweeping dissociation between “vision-for-perception” and “vision-for-action,” it is once again entirely consistent with the intuitive picture of visuomotor action encapsulated in the Control Thesis: Motor systems draw on a combination of conscious and non-conscious sources of visual information in programming visually guided behaviors, and the extent of the role played by conscious visual experience in visuomotor programming depends very much on the type of action the subject undertakes to perform and the constraints on its performance.

It should also be mentioned at this juncture that even when action is performed under ideal conditions for dorsal control, there is evidence that consciously encoded spatial information will make measurable contributions to motor programming. Philosophers, when citing the famous 1995 study conducted by Aglioti et al. (1995), sometimes report that illusory size information has *no* influence on grasping. For example, Clark claims that “mistaken estimates of relative size [in the Ebbinghaus illusion] do not ... affect subjects’ abilities (in the physical, poker chip version) to form precision grips that perfectly anticipate the true size of the centre disks” (Clark, 2009, p. 1461). But this is not quite right. The study by Aglioti et al. (1995), for instance, found that the illusion had a 2.5 mm effect on conscious perception and a 1.6 mm effect on action, as measured, respectively, by the opening between index finger and thumb in a manual estimate of disk size and MGA.¹² Once again, this suggests that conscious visual perception plays *a* role in programming the spatial parameters of visually guided behavior, even as the divergence reveals that other sources of visual spatial information are also in play.

3.4. *Is visual experience allocentric?*

In this section, we address a final challenge to the Control Thesis. According to proponents of TVSH, visual experience makes use of an allocentric or scene-relative spatial coding system as opposed to an egocentric or viewer-relative spatial coding system (Milner & Goodale, 1995/2006, pp. 25–66). Visuomotor control, however, requires information about the egocentric or agent-relative spatial properties of perceived objects; for example, to move one’s hand and take a hold of a hammer, it is necessary to represent the location of its handle relative to the current location of one’s own body (or some part of one’s own body). The supposed allocentricity of visual experience thus undermines the claim that its contents are suitable for playing a direct role in guiding our motor engagements with the world. In support of this argument, proponents of TVSH have put forward two arguments, which we will summarize and then critique. The first, which we discuss in Section 3.4.1, appeals to perceptual constancy phenomena. The second, which we discuss in Section 3.4.2, appeals to evidence, discussed above, that visuospatial awareness is more sensitive than visuomotor action to contextual or “pictorial” depth cues in the stimulus array.

3.4.1. *Visual constancy does not require exclusively allocentric representations*

We shall begin with the argument from perceptual constancy, encapsulated in the following passage:

In sharp contrast to the viewer-based coding required for visuomotor control, visual coding for the purposes of perception must deliver the identity of the object independent of any particular viewpoint . . . [T]he essential problem for the [ventral] perceptual system is to code (and later recover) object identity—thus approximating what Marr (1982) called an “object-centered” description. It is objects, not object views, that the perceptual system is ultimately designed to deliver. As a consequence, human perception is characterized by “constancies” of shape, size, color, lightness, and location, so that the enduring characteristics of objects can be maintained across different viewing conditions. (Milner & Goodale, 1995/2006, pp. 41–42; also see p. 165)

The idea here is that as the ventral stream is responsible for representing the viewpoint-independent properties of the visual world, its outputs are unsuitable for use in motor programming, which requires information about objects’ egocentric locations. In response to this argument, we will note several things. First, the argument is premised on two questionable assumptions about perceptual representations of viewpoint-invariant or “enduring” properties of perceived objects: that the spatial contents of such representations are necessarily unsuitable for motor programming, and that such representations cannot be coded in a “viewer-based” spatial framework. The problem with the first assumption is that, as Milner and Goodale themselves argue, the representation of certain spatial constancies is actually necessary to guide action effectively. In particular, the dorsal stream must compute the intrinsic size of a target object (and not just its relative size)

to control grasping: “It is not enough to know that the target object is larger or smaller than neighboring objects; the visuomotor systems controlling hand aperture must compute its absolute size” (Milner & Goodale, 1995/2006, p. 170). They also accept that action-guiding systems in the dorsal stream are dependent on information about an object’s intrinsic 3D shape (Milner & Goodale, 1995/2006, p. 218; for discussion see Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997). We agree with these claims, and the crucial point here is that if such representations in the dorsal stream have a spatial content that suits them to contribute to motor programming, then representations of absolute shape and size in the ventral stream should also be able to play a similar role.

The problem with the second assumption is that it is clearly possible to represent an object’s visible, intrinsic properties by means of a viewer-based or egocentric spatial coding system. (Indeed, Milner and Goodale claim that action-guiding systems in the dorsal stream represent intrinsic size and visible surface geometry in egocentric terms.) An object’s intrinsic size, for example, can be computed on the basis of its perceived distance from the viewer and its perceived angular size, the latter corresponding to the apparent difference in direction from the eyes between two points on the object’s occluding contour (McCready, 1985); and similarly, an object’s intrinsic shape can be recovered from information about the *slant* and *tilt* of points on its facing surfaces available in what David Marr called the 2.5D sketch (Marr, 1982, chap. 4).¹³ In short, representation of intrinsic, viewpoint-invariant spatial properties does not seem to require use of an allocentric or “scene-based” spatial framework of a sort that is unsuited for direct visuomotor guidance.

However, the more fundamental problem with this entire line of argument is that it rests on a mistaken conception of what it means for the “enduring characteristics” of things to be represented in visual experience. While it may be the case that high-level areas in inferotemporal cortex are designed to deliver allocentric or “object-centered” representations of various kinds—Marr (1982) referred to these as 3D models—there is little reason to suppose that constructing these representations is *the* “essential problem” for ventral areas that most directly support conscious vision, since as Ray Jackendoff has pointed out, 3D models are at once too detailed and too abstract to correspond to the spatial contents of visual experience. They are *too detailed*, because a 3D model “includes the whole of a viewed object, and it does not distinguish between ‘seen’ and ‘unseen’ portions” (Jackendoff, 1987, p. 293), whereas visual experience is by its nature *perspectival*: when you see an object, you do not see the whole thing at once (its complete volumetric structure), but only its facing sides. At the same time, 3D models are also *too abstract* because they do not include information about objects’ distance, direction, and orientation in space. This is precisely why psychophysical and theoretical research in perceptual psychology has standardly assumed that visual experience represents viewer-relative spatial layout by means of something akin to Marr’s 2.5D sketch—a representation of the way visible surfaces are arrayed and oriented in depth relative to the perceiver’s eyes (for discussion, see Nakayama, He, & Shimojo, 1995; Fleming & Anderson, 2004). This has also been the prevailing view in the philosophy of perception (Burge, 2010; Evans, 1982; Peacocke, 1992). When you visually experience an object, your experience

represents it as spatially related *to you*, as opposed to providing you with a detached “view from nowhere” (for discussion, see Schwenkler, 2014). And in representing the positions of surrounding objects relative to your own body, visual experience thus includes the egocentric content necessary to contribute directly to the guidance of action.

Put differently, the problem with the argument from perceptual constancy is that it assumes that visual experience cannot present us with spatially constant “objects” and *also* with inconstant “object views,” when in fact its representational content includes both of these things at once. For example, consider how the intrinsic spatial property of 3D surface curvature is visually represented. As you move in relation to a surface, say that of a disk, the viewer-relative distances, directions, and orientations of points on its surface change systematically. The important point is that you visually experience the disk as invariantly flat across changes in perspective not *in spite of* this variation in your spatial relations to its surface, but rather *because* of it: The “enduring characteristics” of the surface are revealed precisely in the changing relations it bears to you as you move. Similarly, to experience the shape of the disk in a viewer-relative manner is not incompatible with experiencing its “objective” characteristics: You see the disk as intrinsically circular even when viewing it from an oblique angle precisely *because* you see the different egocentric distances and directions instanced by points on its rim. Last, consider the phenomenon of size constancy. What is it to experience the disk’s intrinsic size—as opposed to its angular size or its size relative to other objects in the visual field? One influential answer, going back to Malebranche in the early 18th century, is that an object’s apparent size is scaled to the perceived size of one’s own body (see Simmons, 2003). Evidence for this view comes from studies that have found that an object’s apparent size varies when subjects are made to experience ownership of either a very large (400 cm) or very small (30 cm) artificial body (van der Hoort, Guterstam, & Ehrsson, 2011). In all these ways, the perceiver-relativity of the spatial content of vision is not incompatible with its representing the objective properties of things; to the contrary, vision represents such properties precisely *in a* perceiver-relative way. While the remarks in this paragraph deserve elaboration beyond the space constraints of this article (see Briscoe, 2008b; Burge, 2010; Rescorla, 2014; Smith, 2002 for further discussion), they hopefully serve to motivate the suggestion that perceptual constancy phenomena do not in themselves reveal that visual experience has an exclusively allocentric spatial content.

3.4.2. *Studies of visual illusion do not show that visual experience is exclusively allocentric*

The argument from visual illusions has attracted more attention in the literature, but it is equally flawed. As we showed in Section 3.3 above, numerous studies have reported that geometrical visual illusions have a greater effect on visual experience than on visually guided action; for example, in the Ebbinghaus illusion, the contextual factors introduced by surrounding scene elements significantly influence the target disk’s consciously perceived size, but they have a comparatively smaller effect on the way the subject reaches to grasp the disk. As we noted above, Milner and Goodale interpret this evidence as showing that visual experience represents the spatial properties of perceived

objects using a scene-relative and not a viewer-relative spatial frame of reference (1995/2006, chap. 6).

However, this argument conflates evidence that contextual stimulus factors can *influence* how we visually experience an object's spatial properties with evidence that conscious vision makes (exclusive) use of an allocentric or scene-relative *spatial frame of reference*. This is a conflation because there is evidence that conscious vision makes use of both contextual and local stimulus information when estimating an object's viewer-relative, egocentric spatial properties. Distance in depth, for example, is a paradigmatically egocentric spatial property. But an object's perceived distance from the eyes is known to depend on contextual stimulus information, in particular "pictorial" cues such as texture gradients, shading, relative position, reflections, etc. Perceived depth, in turn, can influence perceived angular size, another egocentric spatial property (Murray, Boyaci, & Kersten, 2006). A fully developed discussion of these matters would involve careful assessment of the evidence concerning the various contextual stimulus factors that contribute to different visual illusions. However, the main point for present purposes is that mere *sensitivity* to contextual information in visual object perception does not entail that visual experience represents *only* the allocentric spatial properties of perceived objects. Hence, it does not speak against the assumption that at least some of the spatial information present in visual experience can be drawn on in programming visuomotor actions.

4. Conclusion

Our aim in this study has been to argue that the commonsense conception of visuomotor control encapsulated in the Control Thesis is also supported by experimental evidence concerning the respective contributions of conscious and non-conscious representations to visuomotor control. To this end, our argument has been as follows:

- As we argued in Section 3.1, the Control Thesis says that in certain circumstances spatial information present in conscious vision is *necessary* for normal visuomotor control, but not *sufficient*, as it needs to be integrated with proprioceptive information concerning the spatial configuration of the body. The thesis is thus compatible with the coexistence of intact conscious vision and deficient visuomotor control in optic ataxia. Moreover, optic ataxia seems to involve a breakdown in systems that monitor hand position and/or integrate proprioceptive hand position signals with visual target information (Blangero et al., 2007; Granek et al., 2012).
- As we argued in Section 3.2, patient DF's intact visuomotor abilities are limited to relatively simple behaviors and break down in the case of behaviors that are more complex. This divergence is compatible with the Control Thesis, which allows that conscious vision is not necessary for actions that can be performed in a routine or automated manner. In contrast, it conflicts with the predictions most naturally drawn from the P-A Model. This evidence suggests a role for ventral area LO,

which is damaged in DF, in supplying viewer-relative information about target shapes that is used in programming motor instructions.

- Similarly, as we argued in Section 3.3, the pattern of perception/action divergences observed in illusion studies is just as the Control Thesis would have it; in particular, visually guided behaviors are more subject to conscious visual illusions to the extent that these behaviors are awkward, challenging, or unpracticed.
- Finally, as we argued in Section 3.4, there is no good reason to believe that the spatial content of conscious visual experience is exclusively allocentric in a way that renders it unfit for direct visuomotor control. Instead, this content includes *both* object-centered *and* perceiver-relative characteristics of perceived objects, and as such it is suited to contribute spatial information that is apt for use by motor systems.

These considerations indicate that there is no general, empirically based objection to the notion that consciously encoded information is used to control bodily actions in the manner suggested by the Control Thesis, and indeed that this thesis is better supported by the available empirical evidence than the P-A Model of visuomotor control. They suggest that even though, in certain conditions, action-guiding visuomotor networks are able to operate without significant conscious involvement, there are various other conditions, far from exceptional, under which systems supporting visual awareness make direct contributions to motor programming—that is, contributions beyond those of selecting targets and types of actions to perform, and supplying semantic information.

In contrast, in rejecting the Control Thesis, the P-A Model posits a “monkey see, zombie do” picture of the relationship between conscious vision and visually guided action: Visual experience enables selection of targets to engage and types of actions to perform, but the egocentric spatial information used to program those actions is computed independently by non-conscious systems in the dorsal stream. The empirical evidence reviewed here conflicts with this view, and thus supports a far less revisionary account of the role of conscious seeing in everyday bodily action. Consciously encoded spatial information seems to be utilized in programming visuomotor action in exactly the cases where the Control Thesis predicts it should. It is the Control Thesis and not the P-A Model that gives us the best interpretation of the contributions to the guidance of behavior made by conscious and non-conscious vision, respectively.

To be sure, the case of DF reveals the surprising extent to which skillful visuomotor action is possible in the absence of conscious seeing. Thus, Goodale and Milner may be right when they remark in a recent paper that “Since these residual visuomotor skills allow DF to live a remarkably full life, they can hardly be considered trivial” (2010, p. 66).¹⁴ But the evidence surveyed above suggests that the contribution of visual consciousness to motor programming may *also* be much less trivial than the P-A Model makes it out to be. It is an empirical question whether continuing to test DF and other subjects on tasks that require novel, difficult, and complex visually guided behaviors (as Milner and Goodale admit we usually have not; see e.g., their 2006, p. 248) will provide additional evidence for the Control Thesis and against the P-A Model, but our prediction is that it will.

Acknowledgments

The authors are listed in alphabetical order. We wish to thank Berit Brogaard, Paula Droege, Markus Schlosser, and Robert Van Gulick for helpful discussions as well as Robert Foley, Melvyn Goodale, Thomas Schenk, Wayne Wu, Silvano Zipoli, and an anonymous referee for detailed comments that led to significant improvements in this study. RB is grateful also to audiences at the University of Pittsburgh and University of Milan and JS to an audience at the 2011 meeting of the Society for Philosophy and Psychology.

Notes

1. For an overview of this distinction, see Goodale and Milner (2004b). Similarly, Jacob and Jeannerod (2003) distinguish between *pragmatic* and *semantic* processing of visual inputs. For further discussion of the planning/programming distinction, see Clark (2007); Dijkerman, McIntosh, Schindler, Nijboer, and Milner (2009), and Goodale and Milner (2010). For a critical appraisal, see Schenk (2010). Though we accept Schenk's observation that the distinction is "somewhat fuzzy" (p. 2828), we will assume for present purposes that it is clear enough to be of theoretical utility.
2. For a challenge to the assumption that conscious oversight interferes with the efficient execution of skilled actions, see Montero (2010).
3. However, as we point in Sections 3.2 and 3.3, there is psychophysical evidence that bottom-up spatial information originating in the ventral stream also contributes to programming well-learned, routine actions when they are performed with the left-hand, or slowly, or under monocular viewing conditions.
4. TVSH also departs from the Ungerleider–Mishkin model with respect to the neuroanatomy of the two processing streams. According to Milner and Goodale, the inferior parietal lobe is not part of the dorsal processing stream but "may turn out to function principally as an elaboration of the ventral stream, by providing perceptual representations of the actions of others as well as the perception of spatial relations among elements in a scene" (1995/2006, pp. 220–221).
5. However, pure optic ataxics with unilateral lesions are in general able accurately to perform hand and arm movements targeted on stationary objects in *central* vision. As perceivers typically foveate visual targets when reaching, optic ataxics do not always notice their visuomotor deficits (Rossetti, Pisella, & Vighetto, 2003; Rossetti et al., 2005).
6. We should emphasize that this includes selecting the action that best achieves *end-state comfort*. Rosenbaum, Heugten, and Caldwell (1996) review findings that a subject will grasp an object with an awkward or uncomfortable initial hand posture if this will result in a more comfortable final posture given what is *intended* to be done with the object. For example, a waiter will typically grasp an inverted, empty glass with his hand in an initial thumb-down position because this will allow him subsequently to hold the glass with his hand in a comfortable and

stable, thumb-up position when pouring water into it and when placing it on the table (Rosenbaum, Cohen, Meulenbroek, & Vaughan, 2006, p. 11). Thanks to Melvyn Goodale for stressing the need to make this point.

7. For a challenge to the claim that dorsal stream representations do not contribute to visual consciousness, see Wu (2014).
8. A recent imaging study by Bridge et al. (2013), however, found thinning of the gray matter in posterior regions of the intraparietal sulcus (IPS) in both hemispheres in DF. Melvyn Goodale has pointed out to us that it is thus possible that DF's poor performance for targets in the visual periphery may be due to damage in her dorsal stream.
9. Thanks to Robert Foster for pushing us to clarify this point.
10. This challenge is further supported by a study carried out by McIntosh, Dijkerman, Mon-Williams, and Milner (2004), who found that DF was impaired in inserting her fingers into holes cut into transparent rectangles, whether those holes were presented as the edges of a single opaque rectangular object, as gaps in the top and bottom of a single "H"-shaped object, or as holes at the centers of two separate squares. In contrast, DF was not impaired in grasping pairs of rectangular objects separated by distances of 30, 40, or 50 mm. As the authors note, these findings challenge the idea that DF's performance in the Dijkerman et al. (1998) study was due to impairments in coding spatial separation, as such information was required for her to grasp the rectangles in pairs. They go on to suggest, in line with our observation below, that the "stringent accuracy requirements" of the hole-grasping tasks "force the subject to monitor the whole act of prehension consciously and deliberately" (McIntosh et al., 2004, p. 702)—an ability that would clearly be compromised in an individual with ventral stream lesions.
11. That is, the task fits Norman and Shallice's (1986/2000) categories (iii) and (iv) of tasks that require deliberate attentive resources, namely those that "are ill-learned or contain novel sequences of actions" and "are judged to be dangerous or technically difficult" (see Section 1 above). Though this claim is based on introspection, it could be evaluated experimentally, for example, by measuring the performance of normal subjects in such a task under conditions of distraction or heavy cognitive load. Also relevant is whether grip posture in such a task is susceptible to visual illusions; here see Section 3.3 below.
12. We should also mention that there is ongoing debate both about the etiology of these effects and the proper method for calculating their magnitude. In particular, Franz (2003) and Franz and Gegenfurtner (2008) have argued that manual estimation responds to changes in object size more strongly than the other measures. In particular, the slope of the linear function relating physical size to manual estimation (1.6) is approximately twice as steep as the slope of the function relating physical size to MGA in grasping (0.82). Correcting for these different response functions by dividing the measured illusion effect by the corresponding slope, Franz and Gegenfurtner argue, significantly eliminates the apparent difference in illusion effect between MGA and manual estimation. However, Goodale (2011) has responded that there

are cases that the “difference in slope” argument does not seem able to handle. In particular, it does not seem able to explain why subjects pick up targets disks in the Ebbinghaus illusion that differ in physical size, but that are identical in *apparent size*, using different grip apertures (Aglioti et al., 1995).

13. Indeed, even non-spatial perceptual constancy mechanisms may involve egocentric spatial representations. Fleming and Anderson (2004), for example, review evidence that egocentrically encoded depth assignments play a critical role in both lightness constancy and perception of a surface’s degree of transparency. Whether a surface appears dark or light can depend on its perceived position and orientation in space relative to the viewer.
14. Though as we noted above, many of DF’s intact perceptual and visuomotor abilities may rely on “tricks” that she has picked up over the years; here see Murphy, Racicot, and Goodale (1996) and Lê et al. (2002). Schenk and McIntosh (2010, pp. 54–55) also present evidence which they take to suggest that DF’s visuomotor abilities may rely on non-standard sources. We should also note that the “fullness” of DF’s life is not itself without controversy. For example, Rossetti et al. claim that “patients with pure optic ataxia can perform everyday actions, whereas patients with visual agnosia are strongly impaired in their daily behavior” (Rossetti et al., 2003, p. 172).

References

- Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*, 679–685.
- Bach, K. (1978). A representational theory of action. *Philosophical Studies*, *34*, 361–379.
- Battaglia-Mayer, A., & Caminiti, R. (2002). Optic ataxia as a result of the breakdown of the global tuning fields of parietal neurones. *Brain*, *125*, 225–237.
- Blangero, A., Ota, H., Delporte, L., Revol, P., Vindras, P., Rode, G., Boisson, D., Vighetto, A., Rossetti, Y., & Pisella, L. (2007). Optic ataxia is not only “optic”: Impaired spatial integration of proprioceptive information. *NeuroImage*, *36*, T61–T68.
- Bridge, H., Thomas, O. M., Minini, L., Cavina-Pratesi, C., Milner, A. D., & Parker, A. J. (2013). Structural and functional changes across the visual cortex of a patient with visual form agnosia. *The Journal of Neuroscience*, *33*, 12779–12791.
- Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, *29*, 336–342.
- Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics*, *59*, 456–469.
- Briscoe, R. (2008a). Another look at the two visual systems hypothesis: The argument from illusion studies. *Journal of Consciousness Studies*, *15*, 35–62.
- Briscoe, R. (2008b). Vision, action, and make-perceive. *Mind & Language*, *23*, 457–497.
- Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, *79*, 423–460.
- Briscoe, R. (2015). Cognitive penetration and the reach of phenomenal content. In J. Zeimbekis, & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 174–199). Oxford, England: Oxford University Press.
- Burge, T. (2010). *Origins of objectivity*. Oxford, England: Oxford University Press.

- Buxbaum, L. J., Schwartz, M. F., & Carew, T. G. (1997). The role of semantic memory in object use. *Cognitive Neuropsychology*, *14*, 219–254.
- Byrne, A. (2009). Experience and content. *The Philosophical Quarterly*, *59*, 429–451.
- Campbell, J. J. (2002). *Reference and consciousness*. Oxford, England: Oxford University Press.
- Carey, D. (2001). Do action systems resist visual illusions? *Trends in Cognitive Sciences*, *5*, 109–113.
- Carey, D. P., Dijkerman, H. C., & Milner, A. D. (1998). Perception and action in depth. *Consciousness and Cognition*, *7*, 438–453.
- Carey, D. P., Harvey, M., & Milner, A. D. (1996). Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. *Neuropsychologia*, *34*, 329–337.
- Castiello, U., Paulignan, Y., & Jeannerod, M. (1991). Temporal dissociation of motor responses and subjective awareness a study in normal subjects. *Brain*, *114*, 2639–2655.
- Clark, A. (2001). Visual experience and motor action: Are the bonds too tight? *Philosophical Review*, *110*, 495–519.
- Clark, A. (2007). What reaching teaches: Consciousness, control, and the inner zombie. *British Journal for the Philosophy of Science*, *58*, 563–594.
- Clark, A. (2009). Perception, action, and experience: Unraveling the golden braid. *Neuropsychologia*, *47*, 1460–1468.
- Creem, S. H., & Proffitt, D. R. (2001). Grasping objects by their handles: A necessary interaction between cognition and action. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 218–228.
- Culham, J. C., Danckert, S. L., De Souza, J. F. X., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research*, *153*, 180–189.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). San Diego, CA: Academic Press.
- Daprati, E., & Gentilucci, M. (1997). Grasping an illusion. *Neuropsychologia*, *35*, 1577–1582.
- Dijkerman, H. C., McIntosh, R. D., Schindler, I., Nijboer, T. C. W., & Milner, A. D. (2009). Choosing between alternative wrist postures: Action planning needs perception. *Neuropsychologia*, *47*, 1476–1482.
- Dijkerman, H. C., Milner, A. D., & Carey, D. P. (1998). Grasping spatial relationships: Failure to demonstrate allocentric visual coding in a patient with visual form agnosia. *Consciousness and Cognition*, *7*, 424–437.
- Dijkerman, H. C., Milner, A. D., & Carey, D. P. (1999). Motion parallax enables depth processing for action in a visual form agnostic when binocular vision is unavailable. *Neuropsychologia*, *37*, 1505–1510.
- Evans, G. (1982). *The varieties of reference*. Oxford, England: Clarendon Press.
- Fleming, R., & Anderson, B. (2004). The perceptual organization of depth. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 1284–1299). Cambridge, MA: MIT Press.
- Forde, E. M. E., & Humphreys, G. W. (2000). The role of semantic knowledge and working memory in everyday tasks. *Brain and Cognition*, *44*, 214–252.
- Franz, V. H. (2003). Planning versus online control: Dynamic illusion effects in grasping? *Spatial Vision*, *16*, 211–223.
- Franz, V., & Gegenfurtner, K. (2008). Grasping visual illusions: Consistent data and no dissociation. *Cognitive Neuropsychology*, *25*, 920–950.
- Gonzalez, C., Ganel, T., & Goodale, M. A. (2006). Hemispheric specialization for the visual control of action is independent of handedness. *Journal of Neurophysiology*, *95*, 3496–3501.
- Gonzalez, C. L. R., Ganel, T., Whitwell, R. L., Morrissey, B., & Goodale, M. A. (2008). Practice makes perfect, but only with the right hand: Sensitivity to perceptual illusions with awkward grasps decreases with practice in the right but not the left hand. *Neuropsychologia*, *46*, 624–631.
- Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, *51*, 1567–1587.
- Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67*, 181–207.

- Goodale, M. A., Jakobson, L. S., Milner, A. D., Perrett, D. I., Benson, P. J., & Hietanen, J. K. (1994). The nature and limits of orientation and pattern processing supporting visuomotor control in a visual form agnostic. *Journal of Cognitive Neuroscience*, *6*, 46–56.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.
- Goodale, M. A., & Milner, A. D. (2004a). *Sight unseen*. Oxford, England: Oxford University Press.
- Goodale, M. A., & Milner, A. D. (2004b). Plans for action. *Behavioral and Brain Sciences*, *27*, 37–40.
- Goodale, M. A., & Milner, A. D. (2010). Two visual streams: interconnections do not imply duplication of function. *Cognitive Neuroscience*, *1*, 64–66.
- Goodale, M. A., Pélisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, *320*, 748–750.
- Granek, J. A., Pisella, L., Blangero, A., Rossetti, Y., & Sergio, L. E. (2012). The role of the caudal superior parietal lobule in updating hand location in peripheral vision: Further evidence from optic ataxia. *PLoS ONE*, *7*, e46619.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*, 187–203.
- Hesse, C., Ball, K., & Schenk, T. (2012). Visuomotor performance based on peripheral vision is impaired in the visual form agnostic patient DF. *Neuropsychologia*, *50*, 90–97.
- Hodges, J. R., Bozeat, S., Ralph, M. A. L., Patterson, K., & Spatt, J. (2000). The role of conceptual knowledge in object use evidence from semantic dementia. *Brain*, *123*, 1913–1925.
- van der Hoort, B., Guterstam, A., & Ehrsson, H. H. (2011). Being Barbie: The size of one's own body determines the perceived size of the world. *PLoS ONE*, *6*, e20195.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jackson, S. R., & Shaw, A. (2000). The Ponzo illusion affects grip-force but not grip-aperture scaling during prehension movements. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 418–423.
- Jacob, P., & Jeannerod, M. (2003). *Ways of seeing*. Oxford, England: Oxford University Press.
- Jeannerod, M. (2006). *Motor cognition*. Oxford, England: Oxford University Press.
- Kanwisher, N. (2004). The ventral visual object pathway in humans: Evidence from fMRI. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 1179–1189). Cambridge, MA: MIT Press.
- Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. New York: Roberts & Company.
- Koch, C., & Crick, F. (2001). On the zombie within. *Nature*, *411*, 893.
- Króliczak, G., Heard, P., Goodale, M. A., & Gregory, R. L. (2006). Dissociation of perception and action unmasked by the hollow-face illusion. *Brain Research*, *1080*, 9–16.
- Lê, S., Cardebat, D., Boulanouar, K., Hénaff, M.-A., Michel, F., Milner, D., Dijkerman, C., Puel, M., & Démonet, J.-F. (2002). Seeing, since childhood, without ventral stream: A behavioural study. *Brain*, *125*, 58–74.
- Marotta, J. J., Behrmann, M., & Goodale, M. A. (1997). The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. *Experimental Brain Research*, *116*, 113–121.
- Marotta, J. J., DeSouza, J. F. X., Haffenden, A. M., & Goodale, M. A. (1998). Does a monocularly presented size-contrast illusion influence grip aperture? *Neuropsychologia*, *36*, 491–497.
- Marr, D. (1982). *Vision*. New York: W.H. Freeman & Sons.
- McIntosh, R. D. (2000). Seeing size and weight. *Trends in Cognitive Sciences*, *4*, 442–444.
- McIntosh, R. D., Dijkerman, H. C., Mon-Williams, M., & Milner, A. D. (2004). Grasping what is graspable: Evidence from visual form agnosia. *Cortex*, *40*, 695–702.
- McCready, D. (1985). On size, distance, and visual angle perception. *Perception & Psychophysics*, *37*, 323–334.
- Milner, A. D. (2012). Is visual processing in the dorsal stream accessible to consciousness? *Proceedings of the Royal Society B: Biological Sciences*, *279*, 2289–2298.

- Milner, A. D., & Goodale, M. A. 1995/2006. *The visual brain in action* (2nd ed.). Oxford, England: Oxford University Press.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*, 774–785.
- Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E., & Davidson, D. L. W. (1991). Perception and action in “visual form agnosia.” *Brain*, *114*, 405–428.
- Montero, B. (2010). Does bodily awareness interfere with highly skilled movement? *Inquiry*, *53*, 105–122.
- Mon-Williams, M., McIntosh, R. D., & Milner, A. D. (2001). Monocular and binocular distance cues: Insights from visual form agnosia I (of III). *Experimental Brain Research*, *139*, 127–136.
- Murphy, K. J., Racicot, C. I., & Goodale, M. A. (1996). The use of visuomotor cues as a strategy for making perceptual judgements in a patient with visual form agnosia. *Neuropsychology*, *10*, 396–401.
- Murray, S., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*, 429–434.
- Nakayama, K., He, Z., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Vision: An introduction to cognitive science* (pp. 1–70). Cambridge, MA: MIT Press.
- Norman, D. A., & Shallice, T. 1986/2000. Attention to action: Willed and automatic control of behavior. In M. Gazzaniga (Ed.), *Cognitive neuroscience: A reader* (pp. 376–390). Cambridge, MA: MIT Press.
- Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.
- Pelissier, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb. *Experimental Brain Research*, *62*, 303–311.
- Perenin, M., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. *Brain*, *111*, 643–674.
- Pisella, L., Gréa, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D., & Rossetti, Y. (2000). An “automatic pilot” for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience*, *3*, 729–736.
- Prinz, J. (2012). *The conscious brain*. New York: Oxford University Press.
- Rescorla, M. (2014). Perceptual constancies and perceptual modes of presentation. *Philosophy and Phenomenological Research*, *88*, 468–476.
- Rosenbaum, D. A., Cohen, R. G., Meulenbroek, R. G., & Vaughan, J. (2006). Plans for grasping objects. In M. Latash & F. Lestienne (Eds.), *Motor control and learning* (pp. 9–25). Berlin: Springer.
- Rosenbaum, D. A., Heugten, C. M., & Caldwell, G. E. (1996). From cognition to biomechanics and back: The end-state comfort effect and the middle-is-faster effect. *Acta Psychologica*, *94*, 59–85.
- Rossetti, Y., Pisella, L., & Vighetto, A. (2003). Optic ataxia revisited: Visually guided action versus immediate visuomotor control. *Experimental Brain Research*, *153*, 171–179.
- Rossetti, Y., Revol, P., McIntosh, R., Pisella, L., Rode, G., Danckert, J., Tilikete, C., Dijkerman, H. C., Boisson, D., Vighetto, A., Michel, F., & Milner, A. D. (2005). Visually guided reaching: Bilateral posterior parietal lesions cause a switch from fast visuomotor to slow cognitive control. *Neuropsychologia*, *43*, 162–177.
- Sakata, H., Taira, M., Kusunoki, M., Murata, A., & Tanaka, Y. (1997). The parietal association cortex in depth perception and visual control of hand action. *Trends in Neuroscience*, *20*, 350–357.
- Schenk, T. (2010). Visuomotor robustness is based on integration not segregation. *Vision Research*, *50*, 2627–2632.
- Schenk, T., & McIntosh, R. D. (2010). Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, *1*, 52–62.
- Schwenkler, J. (2014). Vision, self-location, and the phenomenology of the “point of view.” *Nous*, *48*, 137–155.
- Shadmehr, R., & Wise, S. P. (2005). *The computational neurobiology of reaching and pointing*. Cambridge, MA: MIT Press.
- Siegel, S. (2010). *The contents of visual experience*. Oxford, England: Oxford University Press.

- Simmons, A. (2003). Spatial perception from a Cartesian point of view. *Philosophical Topics*, 3, 395–424.
- Smith, A. D. (2002). *The problem of perception*. Cambridge, MA: Harvard University Press.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 548–586). Cambridge, MA: The MIT Press.
- Wallhagen, M. (2007). Consciousness and action: Does cognitive science support (mild) epiphenomenalism? *British Journal for the Philosophy of Science*, 58, 539–561.
- Wann, J. P., Mon-Williams, M., McIntosh, R. D., Smyth, M., & Milner, A. D. (2001). The role of size and binocular information in guiding reaching: Insights from virtual reality and visual form agnosia III (of III). *Experimental Brain Research*, 139, 143–150.
- Wu, W. (2013). The case for zombie agency. *Mind*, 122, 217–230.
- Wu, W. (2014). Against division: Consciousness, information and the visual streams. *Mind & Language*, 29, 383–406.