"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 200 — #1

 \oplus

 \oplus

 \oplus

 \oplus

8

 \oplus

 \oplus

Cognitive Penetration of the Dorsal Visual Stream?

Brad Mahon and Wayne Wu

1 Introduction

The thesis of the cognitive penetration of vision asserts the existence of specific interactions between cognition and vision, namely a type of informational exchange between them. It remains unclear, however, whether cognition affects vision in this way, and even what would count as evidence for cognitive penetration. The issue engages both empirical and philosophical approaches, for on the one hand, the concept of cognitive penetration and the conditions that suffice for it must be clarified, and on the other hand, the claim of cognitive penetration is an empirical claim about interactions within the human mind and brain. As such, progress on the issue requires joint philosophical and empirical exploration in three steps: (1) philosophical clarification of the concepts needed to state the thesis of cognitive penetration; (2) articulation of a precise computational model of the visual system in question and of the informational exchange that satisfies the thesis of cognitive penetration; and (3) experimental work specifically to test such models. We believe that any convincing argument for cognitive penetration requires taking all three steps, a high hurdle. In this chapter, we focus on the first two steps.

Is visually guided action cognitively penetrated? Specifically, we focus on the possible penetration of dorsal visual stream computations by semantic/conceptual representations of the function and purpose of the use of objects. Our goal will be to provide a clear definition of cognitive penetration that is conceptually fruitful and empirically tractable. This is necessary for the formulation of an adequate argument for cognitive penetration. We then provide a computational characterization of the dorsal stream's role in guiding action. In Section 2, we explicate the concept of cognitive penetration. Section 3 then discusses the anatomical structure of the dorsal visual stream, while in Section 4 we examine its functional role in guiding motor action. In Section 5, we present a prima facie case that concepts, hence cognition, penetrate dorsal visual stream computations.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 201 — #2

THE DORSAL VISUAL STREAM 201

 \oplus

 \oplus

 \oplus

2 Conceptual Issues

In this section we shall lay some conceptual groundwork for our discussion, focusing on the notions of modularity, informational encapsulation, semantic information, and concepts. These are theory-laden and controversial issues. Our aim is to be clear on how we shall understand the terms and to use them to raise specific questions about the visuomotor system. Nevertheless, we believe that our understanding of these notions is general enough to garner reasonably widespread acceptance.

We can think of discussion of *modules* as highlighting certain *features* of processes or capacities, and in this regard, Jerry Fodor's *Modularity of Mind* (1983) provides a useful starting place for thinking about the properties of modules of interest to any cognitive psychologist (for recent empirical discussions of the notion, see Barrett and Kurzban 2006; Mahon and Cantlon 2011). In his list of such properties, Fodor included: information encapsulation, domain specificity, shallow outputs, dedicated neural realization, a characteristic developmental profile, and being fast, automatic, and innate. He never provided, however, a definition of modularity, but in later work he identified informational encapsulation as the 'essence' of modularity (Fodor 2001):

Imagine a computational system with a proprietary...database. Imagine that this device operates to map its characteristic inputs onto its characteristic outputs...and that, in the course of doing so, its informational resources are restricted to what its proprietary database contains. That is, the system is 'encapsulated' with respect to information that is not in its database....That's what I mean by a module. In my view, it's informational encapsulation, however achieved, that's at the heart of modularity. (Fodor 2001: 63)

We follow Fodor and focus on informational encapsulation, so the core issue concerns the exchange of *informational content* between processes and psychological domains. A specific issue that has been much discussed concerns whether the visual domain is informationally encapsulated from the cognitive domain, or whether encapsulation fails and vision is thereby cognitively penetrated by the cognitive domain.

In general, the question is whether two putative modules, X and Y, exchange information in a particular way. More specifically, consider the computations carried out by X that draw on information from its proprietary database and its inputs. We can then say that X is informationally encapsulated from Y if and only if X cannot use information from Y in its computations. Let us think of Y as a cognitive capacity. Accordingly, the failure of informational encapsulation of X relative to Y entails the *cognitive penetrability* of X by Y, to use a phrase of Zenon Pylyshyn's (1984). So, if the visual system is not informationally encapsulated from certain semantic systems, then the visual system is to an extent cognitively penetrated, or so we shall argue. When we speak of failure of informational encapsulation, we intend a version of cognitive penetration.

Failure of informational encapsulation can come in two flavors. On the one hand, it might be the case that Y *can* informationally penetrate X but does not do so under normal circumstances. So, X is *effectively* informationally encapsulated relative to Y.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 202 — #3

 \oplus

202 B. MAHON AND W. WU

This allows that Y could, in exceptional circumstances, informationally penetrate X's computations. Accordingly, we can speak of *potential* informational penetration. This is not the case that interests us. Rather, we are interested in cases of *actual* informational penetration of X by Y, namely where the informational resources from Y are drawn on by X for use in X's computation. Should this occur, then there is a further question as to the scope of this informational interaction: how pervasive is Y's influence?¹ The informational interaction is the crucial notion, and it is here that precise models of X's computational properties are required. Let us say that X is involved in transformation of certain representations in its proprietary database, including its inputs, so as to compute a certain output type O (e.g. object categorization). Y then informationally penetrates X where the representations of Y are not part of X's proprietary database and yet are used by X in computing O. One of the challenges in the debate about cognitive penetration is that more often than not, the computational role of X is underspecified; accordingly, concrete models of the penetration of X by Y remain obscure and hence hard to assess.

Again, where Y is a cognitive capacity or process that informationally penetrates X, then Y *cognitively* (informationally) penetrates X. What makes Y relevantly cognitive? One response would be to say that where the information in Y is *conceptual*, then Y is cognitive, trading on the philosophical connection between concepts and thoughts: concepts are the building blocks of thought. Unfortunately, the notion of concepts is a matter of some heated controversy both in cognitive science and in philosophy, and so clarity is not achieved by merely invoking it. To make both conceptual and empirical progress on this issue, we propose to identify the conceptual with the *semantic*, as spoken of in some areas of psychology. Thus, we will have cognitive penetration of X by Y when Y encodes semantic representations or content over which X computes.

The notion of semantic information is itself a matter of some controversy, but since our aim is to be as clear as possible about questions and proposals, the relevant sense of semantics at issue will always be task-relative. That is to say, we eschew general definitions of what semantics amounts to in respect of psychological processes, and focus instead on the information that is necessary for performing specific tasks. Specifically, in certain well-defined tasks, the relevant information needed for task performance will be operationally identified as semantic. In experimental cognitive science, 'semantics' is generally operationalized as the information that mediates

¹ Informational encapsulation seems to be logically independent of the other conceptions of modularity that have been promoted by psychologists. For example, if we understand domain specificity to be the restriction of a capacity or process X to specific informational contents, then it seems that whether or not X and Y are informationally encapsulated relative to each other is largely independent of whether or not X and Y are domain-specific. There are, of course, special cases. For example, Y might involve computations over some informational content type A and also informationally penetrate the computations of X. Should X otherwise normally compute over informational content type B where B is in some relevant way distinct from information of type A, then informational penetration might undercut the domain specificity of X for B. Of course, should both X and Y operate only over A, then the failure of informational encapsulation does not undercut the domain specificity of X or Y. Spelling out the logical relations between the properties in Fodor's list is a worthwhile exercise that we shall set aside.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 203 — #4

THE DORSAL VISUAL STREAM 203

 \oplus

 \oplus

 \oplus

 \oplus

the mapping from input to output systems, where input and output systems are modality-specific in terms of both their content and their format (see Caramazza et al. 1990). The clearest demonstration of this approach to defining 'semantic'level information comes from cognitive neuropsychological studies of patients with brain damage. The general criterion for positing a semantic-level impairment in such patients is a demonstration that the patient is impaired across multiple modalities of input and multiple modalities of output. So, for instance, while patients with visual agnosia would be impaired for picture naming (name a picture of a hammer), they would not be impaired for naming 'hammer' when hearing the sound of a hammer pounding, or touching a hammer, or when presented with a dictionary definition of a hammer. Thus, in the case of visual agnosia, the patient does not have a semanticlevel impairment (namely about hammers) but rather an impairment at the level of visual representations that interface with semantics. The evidence that semantics is not involved is the patient's normal performance in naming within other modalities (e.g. sound or touch).

A semantic-level impairment would manifest as impaired naming across all of the modalities through which the patient could be presented a stimulus of a hammer. Similarly, while patients with language impairments may be impaired for naming (potentially in writing or speaking or both) pictures or sounds or definitions, they would be able to (for instance) draw a reasonable hammer from memory (modulo drawing abilities), or match a hammer with a nail in a matching task. Patients with semantic-level impairments in respect of hammers would fail those tasks. An interesting task that remains contentious as to whether failure in it is indicative of a semantic-level impairment, and to which we will return below, is object use. Finally, a source of positive evidence that a patient may have a semantic-level impairment would be provided by impaired performance for making judgments about the typical attributes of an item (Is a hammer typically held by one or two hands? Are carpenters or mechanics more likely to use a hammer?). Again, it is important across all such tests that failures of a task are not due to uninteresting reasons. For instance, one might show that a patient who fails at naming across a range of inputs can repeat the word 'hammer' and read the word 'hammer', indicating that the problem is not with articulation. If a patient with a semantic-level impairment fails to be able to draw pictures from memory (draw a hammer), then it is important to show that the patient can copy pictures reasonably, ruling out low-level problems with drawing per se (and so on). The broader point is that by 'semantic' we mean to pick out an empirically defined class of processes or information that mediate the mapping of input to output systems. The nature of those processes and information will come into sharper focus as we frame the processes that are specifically relevant to exploring the hypothesis that dorsal stream object-directed actions are cognitively penetrated by semantic-level information. Specifically, the semantic information on which we shall focus concerns the function and the purpose of the use of objects: what they are for and what the goal is in using them.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 204 — #5

 \oplus

 \oplus

 \oplus

 \oplus

204 B. MAHON AND W. WU

 \oplus

The psychological explication of 'semantic information' is not disconnected from the notion of concepts as discussed by philosophers. How to link concepts to specific information processes in the brain is a controversial topic, but there does seem to be a plausible link between concepts, characterized as the building blocks of thought, and semantic information in the task-relative sense. We assume that the possession of some concept X implies the presence of certain sorts of semantic information. In particular, concepts of *functional objects* imply the presence of semantic information about the function and purpose of the use of those objects. Such semantic information is degraded in patients with semantic dementia, and these semantic deficits are revealed in a variety of tasks that plausibly implicate defective cognition in respect of semantics, specifically the use and function of objects like hammers and spatulas. Patients can be unable to categorize these objects, describe their use, or understand sentences about them.

Indeed, patients with semantic dementia are often unable to use objects appropriately once they have grasped them (Hodges et al. 2000). The action deficits seen in semantic dementia patients, who plausibly have degraded concepts concerning the use and function of certain objects, suggests that those concepts-and by implication associated semantic information-play a role in guiding action. An empirical prediction is that if such information is required, then semantic processing is needed to bring that information to bear in action such that disruption of such processing would disrupt action. While there is more empirical work to be done in this area, there is data suggesting this role for semantic information: drawing on semantic processing that is irrelevant to but concurrent with reaching for an object such as a spatula can disrupt appropriate reaching (Creem and Proffitt 2001). The hypothesis is that semantic processing used in recalling semantically associated words is also normally called upon by the action processing that guides functionally appropriate reaches to manipulable objects. When reaching for such objects, subjects must draw on semantic information about those objects such as representations of their function and the purposes for which they are to be used. Subjects who performed reach-irrelevant semantic processing tended to reach and grasp objects in functionally inappropriate ways (e.g. by the head of a spatula rather than its handle). The central hypothesis to be investigated, then, is that the involvement of concepts in visually guided action amounts to a form of cognitive penetration of visual processing necessary for functionally appropriate action. The empirical upshot is that the informational correlate of such concepts, semantic information, informs such processing.

To return to conceptual issues, let the information in Y be necessary for performance of task T, or at least normally used to perform task T. Then if T is of a certain relevant kind, the information in Y will count as semantic. In this chapter, the relevant task will be functionally appropriate use of artifacts, so the relevant semantic information will concern the function and purpose of such objects. The central question can then be stated generally as follows: does X compute over the semantic information from Y, information identified as semantic relative to some task T? "09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 205 — #6

THE DORSAL VISUAL STREAM 205

Æ

 \oplus

 \oplus

 \oplus

In this chapter, the specific question is this: does the dorsal visual stream compute over semantic information from areas coding information about object use in tasks involving the appropriate action on those objects? If the answer is yes, then those areas cognitively penetrate the dorsal stream. Having philosophically explicated the hypothesis, the next two steps are to articulate a specific computational model of cognitive penetration and then to experimentally test it. In the remainder of this chapter, we focus on the computational articulation, necessarily leaving experiments for another time. We shall articulate hypotheses about the computations carried out by the dorsal stream and the sources of information that it draws on to perform those computations.

3 What Is the Dorsal Stream?

Anatomically, visual information is communicated at the neuronal level to the brain via ganglion cells in the retinas of the eyes. The overwhelming majority of those retinal ganglion cells send axons that synapse in the lateral geniculate nucleus (LGN) of the thalamus (a small-proportion synapse in the mid brain). From the LGN the great majority of forward connections then synapse in V1, which is the earliest level of cortical visual processing, located at the pole of occipital cortex. The classic model of the dorsal vs. ventral stream is that visual information bifurcates at V1, with a dorsal pathway projecting from V1 to dorsal occipital cortex, motion-sensitive area MT, and terminating in posterior parietal cortex, and a ventral pathway projecting from V1 to inferior and lateral temporal-occipital cortex (e.g. Goodale and Milner 1992; Merigan and Maunsell 1993; Ungerleider and Mishkin 1982). It is also the case that the small proportion of retinofugal fibers that synapse in the midbrain project to regions of the dorsal visual pathway (e.g. Lyon et al. 2010).

Computationally, the dorsal pathway analyzes visual information in the service of physically interacting with the world, including getting an effector (e.g. hand) to the right location in space, and if relevant, shaping that effector to a target object (i.e. grasping). Of course, 'visuomotor' actions are not limited to the hands, although manual action has been an important testing ground for the dissociation of the ventral and dorsal streams. Any motor action that is visually guided is a visuomotor action, such as eye movements, walking around furniture, or jumping onto a ledge. The dorsal pathway supports spatial analysis of the location of objects in the service of this broad array of visuomotor actions.

The ventral pathway subserves identification and semantic analysis, and is the major pathway through which semantic information is derived from visual input. Whereas the dorsal pathway is concerned with the true shape, size, and location of objects in the world, representations within the ventral pathway are largely invariant as regards orientation, size, and distance. For this reason, the ventral pathway is sensitive to contextual effects (e.g. that the moon looks large next to the horizon is a contextual effect that would be a quintessential byproduct of how the ventral stream

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 206 — #7

 \oplus

206 B. MAHON AND W. WU

processes visual information). It is this difference between the ventral and dorsal streams that underlies the claim that the ventral stream, but not the dorsal stream, falls prey to some visual illusions, such as the Tichner Illusion. In this illusion, the context in which a central circle is shown affects its perceived size (a circle appears relatively larger if surrounded by smaller circles, and relatively smaller if surrounded by larger circles). Knowing this does not reduce the illusion—a case of the sort that Fodor and Pylyshyn have appealed to in order to demonstrate the cognitive impenetrability of visual experience. However, if the circles consist of poker chips that can be picked up, then the grip aperture of the thumb and finger can be measured, and it can then be tested whether a grasping action falls prey to the illusion. Experiments suggest that it does not (Aglioti et al. 1995; but see Carey 2001; Franz 2001; Smeets and Brenner 2006). This result is surprising if we think that action is guided by what we visually experience. Consequently, the grip should reflect the illusion. On its face, this result raises the possibility that the dorsal stream is immune to (at least this) illusion, and hence encapsulated from the information that gives rise to the illusion.

Our phenomenological experience is generally thought to be aligned with the ventral pathway; it is an open issue as to whether the dorsal visual pathway is able to provide visual information that we can be aware of. However, we shall not be directly concerned with awareness as a way to divide the two streams. Rather, we emphasize the dorsal/ventral distinction at a computational or information-processing level. For example, it may be reasonable to suppose that while both the dorsal and ventral visual pathways are 'driven' by visual input, the dorsal pathway brings fewer assumptions about the world to bear on how it parses that input. Or, while it is not clear that the computations that form the dorsal visual pathway can be engaged in the absence of visual input, ventral visual pathway processes can be engaged by, for instance, visual imagery in the absence of visual input.²

The cleanest demonstrations of the functioning of the dorsal stream would consist of those situations in which visuomotor action is unequivocally 'unguided' by semantic information. A dramatic demonstration of what visuomotor behavior can look like in the absence of semantic input is provided by neuropsychological cases of blind sight. Blindsight patients have lesions to primary visual cortex—they are cortically blind; 'blindsight' refers to the phenomenon that in some patients, they can continue to make actions that avoid obstacles that 'they cannot see' (Goodale and Milner 2004). This is despite the fact that they lack a geniculostriate (i.e. LGN-V1) pathway for the affected regions of the visual field. In one dramatic example of a patient with a confirmed lesion to all of early visual cortex, the patient could walk around obstacles placed in his path (de Gelder et al. 2008). The performance of such patients

² The point being that the dorsal stream may be aligned with certain brain regions; but merely demonstrating the activation of such regions is not evidence of 'dorsal stream' computations. Such computations are defined in terms of the online computation performed over visual input.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 207 — #8

THE DORSAL VISUAL STREAM 207

 \oplus

indicates that non-geniculostriate pathways (retina \rightarrow midbrain \rightarrow extrastriate cortex or retina \rightarrow LGN \rightarrow extrastriate cortex) are sufficient to support the online calculation of at least very coarse actions (e.g. Lyon et al. 2010; Schmid et al. 2010).

Spared action in the context of impaired perception can also be observed in healthy participants. We noted above the empirical claim that manual grasping actions can show immunity, at least under some circumstances, to some visual illusions. In another demonstration of the independence of action from perception, Goodale and colleagues (1986) had participants point to a target dot that was presented in the periphery. Participants were instructed that when the dot appeared, they should saccade to its location and then point to it. Unknown to the subjects, on a proportion of trials, the dot's position would be slightly adjusted during the time when participants' eyes were in flight from the fixation point to the dot's (original) location. Participants did not notice these target jumps, and their eye movements automatically corrected via an intermediary landing spot. However, the finger-pointing movements were entirely accurate and updated to the new location of the target. In a subsequent study with a similar paradigm, Desmurget and colleagues (1999) used transcranial magnetic stimulation (TMS) to disrupt neural processing in posterior parietal cortex during the saccade to the dot location. The authors found that TMS to posterior parietal cortex disrupted the in-flight correction of the pointing movement by the contralateral hand (contralateral to the TMS pulse), while ipsilateral pointing was unaffected by the TMS pulse.

In some ways, therefore, certain dorsal stream computations are more starkly defined than ventral processes, in that the scope of their inputs is limited to online visual information and the scope of their computations is restricted to adjusting the trajectories of the body through space, in real time. In this way, the dorsal stream is one of the strongest candidates for informational encapsulation. The question is whether dorsal stream computations show cognitive penetration by semantic level information about objects. The semantic information that may be doing the penetration can be generally aligned with the ventral stream, but is not limited anatomically to the cortex that subserves the projection from V1 through to anterior temporal cortex. Relevant semantic, or semantically interpreted, information may also be represented by anatomical structures not generally considered part of the ventral pathway proper, such as prefrontal cortex. This does not change the conceptual issues at stake, since the relevant issue is whether dorsal stream processes are cognitively, i.e. semantically, penetrated.³

So what does the dorsal visual pathway know about the world? Take the situation in which a subject at a party is confronted with a beer mug and a wine glass on the

³ If prefrontal regions, which are known to be heavily interconnected with ventral temporal cortex as well as premotor cortex (Rizzolatti and Matelli 2003), are important for interfacing abstract representations of the visual environment with the goals of actions, then it matters whether (putative) penetration of dorsal stream processes is traceable to prefrontal cortex vs. the ventral visual pathway. This is a matter that must be explored further.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 208 — #9

 \oplus

 \oplus

 \oplus

 \oplus

208 B. MAHON AND W. WU

 \oplus

table before her, filled with appropriate liquids. She also has a plate of food and is speaking to people around her. When she reaches for her wine glass, it is a dorsal stream process that calculates the aperture of her grip in flight as well as the velocity profile of that action. However, if the glass were to look slippery and thus require a different hold, that information may not be appreciated by the dorsal visual pathway the dorsal visual pathway would have to be 'told' in one way or another that the glass was slippery. Similarly, if there was a smudge spot on part of the rim that the subject noticed while reaching toward the glass, she might rotate her grasp in flight so a clean part of the rim will come to her lips, without having to adjust her grip once the glass was prehended. Reorientation of the grip to accommodate such information is not a computation that the dorsal visual pathway, at least classically understood, would perform on its own (i.e. over its own propriety database of visual information). Rather, to perform that computation, the dorsal stream must be 'told' the relevant information. We shall return to this, as it may present an important test case of whether the dorsal stream is 'told' this information in a way that constitutes cognitive penetration.

What becomes apparent at this point in the discussion is that we need a theoretically motivated definition of what should count as action vis-à-vis dorsal stream encapsulation. In other words, in the complex action of reaching out, picking up a glass, and taking a sip of wine, which components (or all?) of that complex action are attributed to dorsal stream computations? Identification of those aspects of the action that are attributable to the dorsal stream would then allow us to specifically frame the question of whether those actions are affected by semantic information.

4 What Aspect of Complex Object-Directed Action Is Subserved by the Dorsal Stream?

Consider again our subject at the party who is confronted with a beer mug and a wine glass on the table before her. Now, if there is to be any motor action in this context, she must select both a target and a response. The first challenge to action, then, is that there are two targets and indeed many possible responses. We can speak of the 'behavioral space' available to the subject at this time in terms of the set of input–output (target–response) mappings. Thus, we presume that the subject wants to drink, and so there are two potential targets (beer and wine). On the other hand, our subject could plan to throw one of the objects at a rival, and so it is also clear that for each object there are (at least) two responses: drinking and throwing. Thus far, we have described four available actions, defined by distinct input–output mappings. This yields what we can call a Many–Many Problem, for in normal action, a subject is presented with many inputs and many outputs that identify possible behaviors. The Many–Many Problem is resolved by action when a specific target is identified and acted on in a particular way.

 \oplus

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 209 — #10

THE DORSAL VISUAL STREAM 209

 \oplus

 \oplus

 \oplus

 \oplus

One version of the problem—what we can call the 'deliberative version'—is solved when the subject decides what to do, say to drink some beer. So, her intention represents the action, a specific input–output mapping. Our subject must then produce the action, and it is here that the *visual* system must confront its version of the Many– Many Problem, what we can call the 'non-deliberative' Many–Many Problem. The challenge is that the target of action exemplifies many visual features, and to each 'actionable' feature, a set of motor responses can be made that would instantiate the intended action. There are, after all, many ways to hold a mug to drink from it. So, we can now imagine a more fine-grained behavioral space that maps different ways to implement the action type, drinking from a beer mug. While the deliberative Problem is solved by forming an intention, the non-deliberative Problem is solved by producing an actual action.

Note that solving the non-deliberative Problem is a temporally extended process that extends to the completion of the action. In the action we are considering, ultimately drinking some beer, we can highlight two relevant phases once the action begins: (a) *reaching* for the target to manipulate it and (b) *manipulating* the target once it has been grasped. Thus, the non-deliberative Many–Many Problem can first be divided (at least empirically) into two separate actions. The empirical evidence motivating this distinction comes from patients with brain lesions that selectively compromise different aspects of complex actions, such as reaching out to take a drink from a glass. The reach-to-grasp action itself may be dissociable on empirical and conceptual grounds into grip scaling and targeting the object in visual space (a question we shall return to).

In a series of studies, Goodale, Milner, and colleagues showed that human patients with selective lesions to bilateral lateral occipital regions can have a dense visual agnosia but intact reaching and grasping of objects. In other words, visual information could not be used to derive semantic information from visual input, but was available to guide action. In contrast, patients with lesions to posterior parietal regions can exhibit impairments for targeting objects with a reaching movement and/or scaling their grip aperture appropriately during object grasping—referred to as 'optic ataxia.' Patients with optic ataxia have normal perception and can extract semantic information from objects normally (i.e. their ventral stream is intact; see also Pisella et al. 2006).

An interesting aspect of the behavior of patients with optic ataxia is that they can execute the complex object-associated actions associated with the function of an object, once the object is in hand. In other words, the impairment in optic ataxia concerns the reach-to-grasp component of the action. Stored knowledge—that for instance, hammers are used for pounding and pounding is accomplished with a swinging motion—can be intact in such patients. The other side of this potential dissociation is also observed: lesions to the left inferior lobule are classically associated with object apraxia. Patients with apraxia of object use may be able to reach out and grasp objects fine; the trajectory of their reaching movement and their grip scaling can

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 210 — #11

Æ

 \oplus

 \oplus

210 B. MAHON AND W. WU

be normal, both processes supported by posterior/superior parietal regions, not the inferior parietal lobule. However, once they have the object in hand they are impaired at manipulating the object in the appropriate way to fulfill its function. Thus, they may reach out and grasp the hammer fluidly and without error, but then not be able to pantomime the swinging motion that is associated with using a hammer to pound nails. The failure in these patients is not because they do not appreciate what the object is. In fact, in many cases patients with apraxia of object use can name the same objects they fail to correctly use. Similarly, such patients may be able to retrieve function knowledge—such as the knowledge that hammers are used to pound nails—while still being unable to actually physically manipulate the hammer to accomplish pounding (for reviews of the clinical evidence, see Johnson-Frey 2004; Mahon and Caramazza 2005; Pisella et al. 2006; Rothi et al. 1991).

The double dissociation between reaching and grasping on the one hand (impaired in optic ataxia) and object use (impaired in apraxia) on the other is empirical motivation for the fractionation of a complex action (such as reaching out to take a drink of beer) into component parts. It is clear that the computations underlying reach-tograsp actions are supported by the classically defined dorsal stream. However, it is not at all clear that complex object-associated manipulations could be supported by the dorsal stream. This is because such manipulations are not given by the visual input—they are stored knowledge that must be accessed. For instance, the knowledge of how pliers are used, or how a wrench is used, once the object is in hand is information that is stored. Rothi, Heilman, and colleagues in their influential model of apraxia have analogized those representations to lexical representations of words (e.g. Rothi et al. 1991). The implication then is that, on a conservative analysis, only reach-to-grasp components of actions are dorsal stream computations. The question that we address herein is whether such computations are cognitively penetrated.⁴

5 Is There Cognitive Penetration of Dorsal Stream Computations?

As we saw in the previous sections, dorsal stream computations are geared towards the generation of appropriate motor actions, and in the case of our party example, specifically the reach-to-grasp component of certain motor actions. The question we now raise is whether there is cognitive penetration of dorsal stream computations by semantic information coded outside the dorsal stream, specifically by semantic information concerning the function and use of objects.

Our question of cognitive penetration concerns whether semantic information influences solving the non-deliberative Many–Many Problem focused on the reach-

⁴ An earlier argument in this direction, but phrased in terms of the conceptual content of vision, was given in Wu (2008), where it was argued that concepts are needed to organize attention in order to serve the appropriate use of artifacts.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 211 — #12

THE DORSAL VISUAL STREAM 211

 \oplus

to-grasp component. Specifically, is semantic information needed to influence dorsal stream computations so as to adequately solve this Problem? Whether this is so is ultimately an empirical question, but let us first flesh out the space of possible ways in which dorsal stream computations could interface with semantic information in order to solve the non-deliberative Problem. We can think of three places that semantic information can influence dorsal stream computations: (a) at the level of input into the dorsal stream; (b) at the level of output from the dorsal stream; and (c) at the level of internal computations within the dorsal stream. It is only the third case where the influence would count as cognitive penetration.

Let us consider the first possibility: that semantic information influences the input into the dorsal stream. One model that suggests this is Goodale and Milner's proposal for how semantic information in the ventral stream affects dorsal stream computations. They write (2004: 102):

Although the dorsal and ventral streams have their own sophisticated and specialized languages, they both still retain contact with the more basic language of the retina. Rather like the human operator can instruct the robot via the two-dimensional optical array behind the lens of the robot's camera, so the ventral stream can instruct the dorsal stream via the common retinotopic map in early visual areas.

On this model, it seems that semantic information from the ventral stream is 'fed back' to early visual areas, presumably to identify certain regions in retinotopic maps for further processing by the dorsal stream. The idea would not be that semantic information is literally transposed or communicated to early visual regions, but that the semantic information would increase, for instance, the gain of neurons coding information in a particular region of the visual field that is processing the information about the selected object. Goodale and Milner invoke the metaphor of an operator of a robot pointing to a region in a map of the terrain around the robot to get the robot to head to that location. Analogously, when our subject decides to drink from the mug, their model holds that the ventral stream analyzes the properties of the mug as an artifact used to drink, and perhaps 'highlights' the relevant region in retinotopic maps corresponding to the retinal projection of the mug's handle. In this way, the dorsal stream focuses on information from that region in space. We might think of this as a form of attention that helps to select relevant information for further processing.

Whether this is how semantic information in the ventral stream influences dorsal stream computations—namely via the mediation of early visual areas affecting input into the dorsal stream—is an open empirical question. What we wish to highlight is that this does not seem to exhaust the role that the ventral stream may have in 'telling' the dorsal stream about semantically interpreted or behaviorally relevant aspects of objects. For example, even if part of the role of semantic information is to aid selection of locations as spatial targets for action by highlighting regions of retinotopic maps, this still leaves underspecified what action would be relevant within those regions. For example, as the Many–Many Problem demonstrates, for any given target there

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 212 — #13

 \oplus

 \oplus

 \oplus

 \oplus

212 B. MAHON AND W. WU

 \oplus

are a variety of possible responses that are available. In the case of the handle on the mug, there are many possible ways to grip it, of which only a subset is appropriate to conventions for the object's use. Put another way, given the two components of action in our example, reaching and grasping, we can see Goodale and Milner's model as showing how semantics could aid reaching (i.e. targeting) by highlighting spatial locations. It does not, however, show how the specific grasp is selected by the dorsal stream, something which presumably must also be sensitive to the function of the targeted object and the intended purposes in using that object, and hence to semantic information regarding that object. Accordingly, even if Goodale and Milner's model is correct, semantic information must play a role in selecting appropriate action in response to visual information. This leaves the other two points of influence: outputs and internal computations of the dorsal stream. Again, only the latter would count as cognitive penetration.

Consider the possibility that semantics influences the output of the dorsal stream. A role for semantics in constraining the outputs of dorsal processing is perhaps easiest to conceptualize if we think of the dorsal stream as providing a set of action plans. Put in terms of the Many–Many Problem, we can think of the dorsal stream as constraining the behavioral options, but not to the extent that a single behavioral path is chosen. The constraints expressed over the multiple behavioral options may be *biomechanical* constraints of the body together with the *physical* (i.e. semantically uninterpreted) constraints about the shapes, sizes, and distances of objects. This yields a set of action plans that could be run given these biomechanical and physical constraints. On this model, further selection of the plans produced by the dorsal stream must be implemented to yield a specific input-output mapping.

At this point, post-dorsal stream computation, appropriate action selection (inputoutput mapping) to implement a specific plan might require semantic information, say reaching for the handle of the mug in the right way. Still, even if the output of the dorsal stream is multiple in this way—a possibility that we leave open—one might think that it must still exhibit selectivity in that the initial behavioral space available to the agent is *very* large. It would seem imprudent for the dorsal stream to make its initial selections in the absence of reducing the Many–Many options without guidance by the relevant semantics. Thus, even if the dorsal stream outputs multiple behavioral options that are consistent with the biomechanical and physical constraints of the action context, if the dorsal stream is not outputting *all possible* actions consistent with these constraints, then dorsal stream computations are influenced by semantics. Otherwise, the number of possible actions would be overly large to allow for efficient action selection following dorsal stream computation. After all, qua a *physical* part, a handle can be acted on in many different ways, but there are fewer *appropriate* ways to act on a handle qua functional part.

We saw above that the actual manipulation of the object is typically not something that is mediated by the fast visuomotor processes of the dorsal visual system. However, reach-to-grasp actions are classic dorsal stream computations. The potentially critical "09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 213 — #14

THE DORSAL VISUAL STREAM 213

Æ

 \oplus

 \oplus

 \oplus

test cases, then, are situations when an object is grasped in a way that anticipates its eventual use. Thus, if a hammer is on the table but oriented such that the handle is pointed away, then the appropriate reach-to-grasp movement would not be the easiest or shortest—the shortest would be to pick up the hammer by the head, or to pick it up with a power grasp with the head facing down as opposed to up. But if one were picking up the same hammer to hand to someone else, then it may be functionally appropriate to grasp the hammer by its head (consider the way one grasps a knife to cut vs. to hand to someone). The same situation is presented when picking up a wine glass that has a smudge on it that one wants to avoid with one's lips.

Compelling data relevant to this point are also provided by patients with object form agnosia. Recall that such patients are unimpaired for grasping objects. However, such patients show dramatic impairments for grasping objects in a way that anticipates the correct use of the object. In other words, the reach-to-grasp actions are well-formed actions, and the grip aperture and orientation are well calibrated to the part of the object that is grasped. But a patient such as the well-studied visual object agnosia patient, DF, would be as likely to pick up a hammer upside down or by the head as by the handle in the correct orientation. Similarly, in grasping a wine glass lying on its side, patients like DF would be as likely to pick it up initially with the cup facing up as with it facing down. Such patients would also be impaired for integrating information about the surface properties of an object (such as whether the handle was slippery) into their initial grasp. We have seen above that the dorsal stream has very definite limitations on the types of information that it can represent. The question is: where does the dorsal stream get the information that eventually constrains action output? Does information about the part of the rim of the glass that is smudged, or that the glass looks slippery, constrain action internal to the computations of the dorsal stream?

We think that these types of considerations fit more naturally with the hypothesis that semantic information constrains the dorsal stream's computations, rather than operating only over the inputs and outputs of the dorsal stream. If this hypothesis were true, it would open up the possibility for cognitive penetration of the dorsal stream, via semantic information influencing dorsal stream computations. We suggest the following sufficient condition for cognitive penetration:

Y semantically penetrates X if semantic information encoded in Y is directly transmitted to X such that X computes over this information to generate its standard output.

Thus, we can think of Y as any non-dorsal stream region that encodes action-relevant semantic information, and X as the dorsal stream. Specifically, if the function of the dorsal stream is to carry out computations that lead to reach-to-grasp actions of the right sort, then we can think of this as a computation that reduces the behavioral options presented in the Many–Many Problem. One way to think of this is that the dorsal stream selects a behavioral path through the available behavioral space, and the role of semantics is to aid in this selection, in *addition* to the standing biomechanical

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 214 — #15

 \oplus

 \oplus

 \oplus

 \oplus

214 B. MAHON AND W. WU

 \oplus

and physical constraints. Dorsal stream computations would be, then, by hypothesis, sensitive to semantic information: they select an input–output map (or mappings) given relevant semantic information. Specifically, the reach and grasp identified by the dorsal stream is appropriate to the proper use of the object in question, say a beer mug.

We can imagine the semantic influence on the dorsal stream in two ways, roughly via specification of the goal and via more direct interactions with the ventral stream. Let us begin with the influence of goals on action selection. Again, we emphasize that what counts as the correct model remains an open empirical question. Our goal is to delineate, at least partly, the relevant space of possibilities. In the case of the influence of goals on action selection, recall the deliberative and non-deliberative Many-Many Problems. We can think of the deliberative Problem, one that is solved by making a decision or forming an intention, as specifying the end point of action. At one level, this specification is abstract: for our subject at the party, it is to drink beer from a specific mug. We can imagine the influence of the representation of a mug invoked in the intention as a semantic informational constraint that affects action-relevant computations. It might delineate relevant action plans by the motor system such as specifying the *endpoint* of action, as it is the endpoint of the action that directly 'interfaces' with the intention. For instance, it is the final position that one wants the hammer to be in after grasping it that will depend on whether one intends to use it or hand it to someone else. The relevant aspects of the motor system are guided by the goals set by intention: if the goal is to drink from that mug, then the hand should be shaped on the handle in such-and-such way, namely in a drinking-appropriate grip. This can then provide a target for the dorsal stream: if the hand should land on this handle in such-and-such way, the dorsal stream must select a trajectory and grip that will achieve this endpoint, given the current layout of the world, the starting position of the hand, and relevant biomechanical constraints. In this way, the solution to the non-deliberative Many-Many Problem is constrained by semantics, and provides a way for semantics to constrain dorsal stream computations (for arguments that intentions do cognitively penetrate visual computations in the parietal cortex, specifically in the case of visual spatial constancy, see Wu 2013).

A second way that semantic information could influence the dorsal stream is via the ventral stream. The relevance of the ventral stream to cognitive penetration is illustrated via the link between semantic dementia and semantic information encoded in the ventral stream. Early in semantic dementia, there are changes in the anterior medial temporal lobe (in addition to the frontal and parietal lobe). As we noted earlier, the symptoms observed in semantic dementia affect cognition of the function of artifacts, and it is the underlying semantic information that is relevant as a source of influence on the dorsal stream. For instance, consider the situation in which the mug must be grasped by its handle in a particular orientation in order to be able to drink from the mug. The identification of the handle (as such) is a process that would be mediated by ventral and not dorsal visual analyses. We have discussed how one "09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 215 — #16

THE DORSAL VISUAL STREAM 215

 \oplus

 \oplus

 \oplus

 \oplus

way in which the dorsal stream could be cued to the handle as the target of the action would be by 'highlighting' the region of space (represented in the early visual cortex) that corresponds to the handle. It is not clear, however, that the notion of a 'handle' as a behaviorally relevant affordance is something that can survive translation into the decidedly 'asemantic' language of the retina. If it cannot, then the expectation would be that the ventral representation of the handle (as such) is communicated directly to dorsal stream computations, such that the trajectory and grip scaling necessary to bring the hand to the handle in the right way can be computed given the physical and biomechanical information that the dorsal stream has available to it.

It is important to note that the anatomical and functional connectivity that these hypotheses would suggest is at least plausibly demonstrated; or rather, there certainly are no data that rule out such connectivity, and there are data that could be taken to indicate the kinds of connectivity that would be expected according to those hypotheses. Specifically, it is known that the ventral temporal cortex is heavily interconnected with the lateral prefrontal cortex, which is in turn heavily interconnected with premotor cortex, which is interconnected with anterior intraparietal sulcus, aIPS (Rizzolatti and Matelli 2003). Thus, the pathway from semantics and intentions to motor planning stages, and the region of parietal cortex known to be critical for grip scaling (aIPS; e.g. Binkofski et al. 1999) is well established. There is also known to be connectivity, both functional (Mahon et al. 2007; 2013; Noppeney et al. 2006) as well as anatomical (e.g. Rushworth et al. 2006) between the relevant regions of temporal and parietal cortex, thus providing a reasonable pathway for the influence of ventral stream information on dorsal stream computations. Further work will be needed to delineate the relevant anatomical connectivity between temporal and parietal cortex as well as overlaying this on work demonstrating functional connectivity.

Overall, we think there is a good prima facie case to be made for semantic influence on dorsal stream computations needed to specify reach-to-grasp movements. Such computations involve constraining behavioral possibilities in light of the nondeliberative Many–Many Problem. Should semantic information, whether mediated by the subject's goals or via semantic information in the ventral stream, be brought to the dorsal stream to aid its solving the Many–Many Problem, this would count as an important and striking case of the failure of informational encapsulation of part of the visual system. Accordingly, there would be one sense in which the visual system, or at least the dorsal stream, is not encapsulated with respect to semantic information about object function and the purposes to which the object will be put to use. As we noted earlier, as such semantic information and, given the tie between concepts and thought, to the failure of encapsulation of the dorsal stream from thought. That is, there is a prima facie case to be made that the dorsal stream is cognitively penetrated.

Given our earlier point that the existence of cognitive penetration will likely be established with empirical inferences that result from careful experiments, there is no philosophical argument that establishes the existence of dorsal stream penetration "09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 216 — #17

 \oplus

 \oplus

216 B. MAHON AND W. WU

by semantic information. Rather, the role of philosophical analysis has been to make clear the possibilities and to formulate the conditions for cognitive penetration. Computational models then provide a picture of what the mechanism of penetration might be, at least at an abstract level. To that extent, and at least in visually guided action, we are closer to assessing the actuality of the cognitive penetration of vision. The next, critical steps are to further elaborate the underlying models and to carry out experiments to test them. We believe that the fruitfulness of cognitive penetration as a matter for psychology and for much of philosophy of mind depends on researchers carrying out such experiments on the basis of the type of analysis that we have undertaken here. Cognition might very well penetrate vision. We hope to have taken the first two steps (of three) to establishing this.

References

- Aglioti, S., DeSouza, J. F. X., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology* 5: 679–85.
- Barrett, H. C., and Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psycholog-ical Review* 113: 628–47.
- Binkofski, F., Dohle, C., Posse, S., et al. (1998). Human anterior intraparietal area subserves prehension: a combined lesion and functional MRI activation study. *Neurology* 50: 1253–9.
- Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cognitive Neuropsychology* 7: 161–89.
- Carey, D. P. (2001). Do action systems resist visual illusions? *Trends in Cognitive Sciences* 5(3): 109–13.
- Creem, S. H., and 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: 218–28.
- De Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., van den Stock, J., Stienen, B. M. C., Weiskrantz, L., and Pegnas, A. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current Biology* 18: R1128.
- Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., and Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nature Neuroscience* 2: 563–67.
- Fodor, Jerry A. (1983). The Modularity of Mind. Cambridge, Mass.: MIT Press.
- Fodor, Jerry (2001). The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology. Cambridge, Mass.: MIT Press.
- Franz, V. H. (2001). Action does not resist visual illusions. Trends in Cognitive Sciences 5: 457-59.
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences* 15: 20–25.
- Goodale, M. A., and Milner, A. D. (2004). Sight Unseen: An Exploration of Conscious and Unconscious Vision. Oxford: Oxford University Press.
- Goodale, M. A., Pelisson, D., and Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*. 320: 748–50.

"09-John-Zeimbekis-Ch08-drv" — 2015/3/25 — 12:34 — page 217 — #18

THE DORSAL VISUAL STREAM 217

 \oplus

 \oplus

- Hodges, J. R., Bozeat, S., Lambon Ralph, M. A., Patterson, K., and Spatt, J. (2000). The role of conceptual knowledge in object use: evidence from semantic dementia. *Brain* 123: 1913–25.
- Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences* 8: 71–8.
- Lyon, D. C., Nassi, J. J., and Callaway, E. M. (2010). A disynaptic relay from superior colliculus to dorsal stream visual cortex in macaque monkey. *Neuron* 28: 270–79.
- Mahon, B. Z., and Cantlon, J. F. (2011). The specialization of function: cognitive and neural perspectives. *Cognitive Neuropsychology* 28(3–4): 147–55.
- Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology* (Paris) 102: 59–70.
- Mahon, B. Z., Kumar, N., and Almeida, J. (2013). Spatial frequency tuning reveals interactions between the dorsal and ventral visual systems. *Journal of Cognitive Neuroscience* 25: 862–71.
- Mahon, B. Z., Milleville, S., Negri, G. A. L., Rumiati, R. I., Caramazza, A., and Martin, A. (2007). Action-related properties of objects shape object representations in the ventral stream. *Neuron* 55(3): 507–20.
- Martin, A., Ungerleider, L. G., and Haxby, J. V. (2000). Category specificity and the brain: the sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (ed.), *Higher Cognitive Functions: The New Cognitive Neurosciences*, 1023–36. Cambridge, Mass.: MIT Press.
- Merigan, W. H., and Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience* 16: 369–402.
- Noppeney, U., Price, C. J., Penny, W. D., and Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses. *Cerebral Cortex* 16: 437–45.
- Pisella, L., Binkofski, F., Lasek, K., Toni, I., and Rossetti, Y. (2006). No double dissociation between optic ataxia and visual agnosia: multiple sub-streams for multiple visuo-manual integrations. *Neuropsychologia* 44: 2734–48.
- Pylyshyn, Z. W. (1984). Computation and Cognition. Cambridge, Mass.: MIT Press.
- Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Experimental Brain Research* 153: 146–57.
- Rothi, L. J. G., Ochipa, C., and Heilman, K. M. (1991). A cognitive neuropsychological model of limb praxis. *Cognitive Neuropsychology* 8: 443–58.
- Rushworth, M. F. S., Behrens, T. E. J., and Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. *Cerebral Cortex* 16: 1418–30.
- Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., Ye, F. Q., and Leopold, D. A. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature* 466: 737–7.
- Smeets, J. B. J., and Brenner, E. (2006). 10 years of illusions. *Journal of Experimental Psychology: Human Perception and Performance* 32: 1501–4.
- Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (eds), *Analysis of Visual Behavior*, 549–86. Cambridge, Mass.: MIT Press.
- Wu, W. (2008). Visual attention, conceptual content and doing it right. Mind 117: 1003-33.
- Wu, W. (2013). Visual spatial constancy and modularity: does intention penetrate vision? *Philosophical Studies* 165: 647–69.