

Binding Through the Fovea: A Tale of Perception in the Service of Action

Paul Cisek¹ & Martine Turgeon²

¹Dept. de physiologie
Université de Montréal
C.P. 6128 Succursale Centre-ville
Montréal, Québec H3C 3J7
CANADA
cisekp@magellan.umontreal.ca
<http://www.cisek.org/pavel>

²Psychology Department
McGill University
1205 Dr. Penfield Ave
Montréal, Québec H3A 1B1
CANADA
martine@hebb.psych.mcgill.ca
<http://www.psych.mcgill.ca/labs/auditory/labmembers.html#turgeon>

Copyright (c) Paul Cisek & Martine Turgeon 1999

PSYCHE, 5(34), December 1999
<http://psyche.cs.monash.edu.au/v5/psyche-5-34-cisek.html>

KEYWORDS: attention, audition, binding problem, two visual systems, sensorimotor control.

COMMENTARY ON: A. David Milner & Melvyn A. Goodale. (1995) *The Visual Brain in Action*. (Oxford Psychology Series, No. 27). Oxford: Oxford University Press. xvii + 248pp. ISBN: 0198524080. Price: \$35 pbk.

ABSTRACT: By characterizing the function of the ventral and dorsal visual streams as, respectively, vision-for-perception and vision-for-action, Milner and Goodale (1995) have brought some action onto the perceptual scene. However, with the distinction of the ventral "what" system and the dorsal "how" system, comes a dilemma: How is the operation of the two systems united toward common goals? This is an example of a *binding problem*. We propose that for binding these two systems, no mechanism needs to exist in the brain to bring ventral and dorsal representations together. Coherent behavior

may be accomplished by focusing the two streams upon the same external object through a strategy of spatial selection, using either the fovea or selective attention. This strategy exploits the simple fact that two objects cannot occupy the same place at the same time. Other kinds of binding, like that involving the perceptual binding of the elements of a scene into coherent perceptual units, may be accomplished by exploiting other regularities of the environment, such as the likelihood that two simultaneous sounds were caused by a common event. From an evolutionary perspective, the most effective way to cope with the distant environment is to use different but complementary sensors, both contributing to aspects of identification and action guidance. For example, audition plays a role both in perceptual analysis and in action guidance, suggesting the possibility of segregated processing of auditory information. Thus, the functional distinction between a "what" system and a "how" system may not be limited to the visual modality, but may be a fundamental distinction for behavioral control in general.

1. The "What" and the "How": Vision for Perception and Vision for Action

Based on converging neurophysiological and behavioral data with animals and clinical data with human patients, Milner and Goodale (1995) offer a new functional interpretation of Ungerleider and Mishkin's (1982) proposed distinction between the "what" and "where" visual systems. Ungerleider and Mishkin (1982) suggested that the "ventral" visual stream (geniculostriate pathway projecting to the inferotemporal cortex) subserves object identification, while the "dorsal" stream (projections from the striate cortex and colliculi to the posterior parietal cortex) subserves object localization. Instead, Milner and Goodale (1995; see also Goodale & Milner, 1992) suggest that the function of the dorsal stream is better described as *mediating visually-guided actions*. Thus, they replace Ungerleider and Mishkin's "what" vs. "where" distinction with a distinction between "what" vs. "how".

In Milner and Goodale's proposed functional architecture, the ventral stream operates in an object-centered frame of reference and is phylogenetically more recent than its dorsal counterpart. It extracts the invariant properties of world objects and events (e.g. shape, orientation, and size) which are independent of the changing viewing conditions (e.g. overall illumination and relative position of the observer and the perceived object); hence it provides the animal with the critical information for identifying the objects in its surroundings. The ventral stream is thus responsible for what is typically meant by conscious perception (i.e., seeing something) and recognition of familiar world objects and events. On the other hand, the phylogenetically older dorsal stream operates in a viewer-centered frame of reference and mediates visually-guided behaviors. Such control can operate without the conscious awareness of the perceiver. In summary, while the ventral stream specializes in identifying external objects (vision in the service of perception), the dorsal stream programs and directs visually-guided actions performed upon these objects (vision in the service of action).

The intrusion of viewer-centered coordinates in representational systems would interfere with the computations leading to perceptual constancies, that is, those properties of world objects that remain constant despite the movement of the observer. For instance, it is a fact of physics that the size of an object remains constant despite changes in viewpoint. It is thus adaptive for a biological system to perceive the size of an object as constant by ignoring the size of its retinal projection. Similarly, the intrusion of orientation-invariant representations in visuomotor control would be too rigid to allow for a precise adjustment of the motor programming of an action with respect to the relevant egocentric parameters such as the position of the observer relative to the object. For instance, computations based on size constancy would prevent the use of relative size, which from projective geometry is directly related to the actual distance to the object. The case of size constancy versus relative-size information is an illustration of the usefulness of segregated processing for perception and action.

It is important to acknowledge that the distinction between the ventral and dorsal visual streams is an idealization of a reality which is much more complex. Anatomical, physiological, and behavioral data suggests that the border between the two systems is not crisp (Carey, 1997; [Mattingley, 1999](#)). Nevertheless, the distinction between vision in the service of perception and vision in the service of action has turned out to be extremely useful, as the examples discussed by Milner & Goodale (1995) illustrate. We do not challenge this usefulness. However, we point out that perception is itself also ultimately used in the service of action.

2. The "What" is for the "How": Perception is Also in the Service of Action

Traditional accounts of vision (e.g. Marr, 1982) implicitly assume that the goal of the visual system is to build a coherent representation of the visual scene. However, as Milner and Goodale recognize, to build an internal model of the external world is not the ultimate goal of perceptual systems. Goodale (1996) emphasizes that although the study of perception as the process of building an internal representation of the external scene from the sensory input is implicit in most current theories of perception, the conscious experience of the world that results from such a process is possibly quite recent on the evolutionary time scale: "Vision did not evolve to enable animals to see the world; it evolved to provide distal control of their movements within it. Conscious sight is a relative newcomer on the evolutionary stage" (Goodale, 1996 p. 390).

According to Milner and Goodale, one of the most central questions in modern neuroscience is: "How is sensory information transformed into purposeful acts?" (Milner & Goodale, 1995 p. 202). Though this question forms the backdrop for many research programs, it is rarely addressed in a perspective as global as that used by these authors. The inevitable specialization that has arisen from the rapidly growing body of data and the complexity of the questions related to perceptual and motor-control processes at

different levels of analysis (functional, physiological, anatomical) has driven many to isolate themselves within issues that fall neatly within some well-delineated discipline. The questions asked by Milner and Goodale force us to think about how perception guides our actions and by the same token reminds us that object identification (the "what") is also *ultimately in the service of action* (the "how").

While the ventral pathway serves as a front end to a representational system that permits the formation of goals and the decisions to engage in specific acts without reference to the specific programming planned by the dorsal stream (i.e., deciding what to do without reference to how to do it), the dorsal pathway directs the course of actions planned by the ventral stream with respect to world objects, possibly without the conscious awareness of the perceiver. Milner & Goodale (1995) provide some highly compelling arguments for the ecological validity of such a functional specialization of the two pathways, based upon an extensive amount of empirical evidence from both the animal and human literature. Notwithstanding its ecological validity, there is another critical aspect inherent to the segregation of these two pathways: their necessary cooperation. As Milner and Goodale recognize, the "what" and "how" ultimately have to work together: "...efficiently programmed and coordinated behaviour requires that neither the ventral nor the dorsal stream work in isolation: they should cooperate" (p. 202).

3. Binding the "What" and the "How"

Brain theorists often discuss what has come to be known as the "binding problem". Neurophysiological experiments have shown that different, and often very distant brain systems specialize in different aspects of perceptual analysis and behavioral guidance. Given such divergence of information, how is this information ultimately integrated into a coherent unambiguous whole? For example, how do neurons involved in representing shape and neurons involved in representing position integrate their independent representations into a coherent representation of shape-at-a-position (von der Malsburg, 1996)? The solution cannot be that every combination of lower-level signals converges onto a neuron dedicated toward identifying that specific combination, because there are far too many possible combinations required (the straw-man of the "grandmother cell"). Instead, there must exist a flexible mechanism for unifying features with other features based on the context in which they appear. A multitude of solutions to the binding problem have been proposed, most of them implicating temporal synchrony (Milner, 1974; Shastri & Ajjanagadde, 1993; Joliot, Ribary, & Llinás, 1994; Llinás, Ribary, Joliot, & Wang, 1994; Singer, 1996; von der Malsburg, 1996; Roelfsema, Engel, König, & Singer, 1997).

Many different kinds of binding problems appear to exist. These may be classified into "within-system" binding and "cross-system" binding. For example, within the system dedicated for object recognition the first question is one of perceptual grouping: How do features belonging to one object get grouped together while they are segregated from the features of other objects (Milner, 1974; Singer, 1996)? How do different aspects (shape,

color, location) of each feature get bound together? An example of a cross-system binding problem is the question of how the visual percept of a cat is connected to the auditory percept of its meowing to yield the cognitive conclusion that food is being requested.

Another example of cross-system binding arises in the context of the dorsal and ventral visual streams. If the ventral stream specializes in identification of objects while the dorsal stream specializes in acting upon them, then how do the two cooperate? How are the "what" and the "how" processes bound? For example, consider an animal attempting to grasp an apple, either with its hand or with its mouth. One part of its visual system, the ventral stream, specializes in identifying objects and contains mechanisms capable of recognizing an apple within the visual array. As described above, to perform this recognition the ventral stream deliberately filters out information on spatial location and produces an orientation- and scale-invariant representation. Meanwhile, the dorsal stream specializes in controlling actions and is able to initiate movements to approach and grasp an object. Its function requires deliberate focus on the egocentric spatial arrangement of objects to be interacted with regardless of their identity. How are these very different representations unified into a coherent whole and used in the cooperative guidance of action?

It is certainly plausible that distributed representations in the cerebral cortex are unified through a general mechanism for solving the binding problem. Perhaps this mechanism involves temporal synchrony, perhaps it involves complicated anatomical connections, perhaps it's something else altogether. However, it is also possible that different kinds of binding problems are resolved with different mechanisms, each dependent upon the requirements of the behavioral task at hand. Furthermore, it is possible that *some of these behavioral tasks do not require independent representations to be bound at all.*

For example, consider the grasping task described above. Does the representation of object-type information *need* to be bound within the brain with the representations of object-arrangement which are used for guiding action? What happens, behaviorally, if it is not?

Let's imagine a creature foraging for food. As it scurries around, moving its body, head, and eyes, different images fall upon its retina and are passed into the visual system. The ventral stream filters the visual information so as to identify features of the viewed objects which define what sorts of actions these objects afford. Some images may be ignored while others are attended. Suppose an apple falls into the visual field and is recognized as something worth eating. Once recognized, the apple captures the attention of the foraging animal, changing its oculomotor pattern from "look around" to "maintain fixation". The decision to grasp the fixated apple can be made by the brain based solely on the ventral stream information. Once the decision to "grasp" is made, the dorsal stream can begin to do its work. *Because the apple is fixated, there is no ambiguity as to what is supposed to be grasped.* The command that releases the appropriate movement mechanisms simply says "grasp whatever is fixated". The action of grasping involves dorsal stream processing which has access to all the information that defines "how" to

grasp regardless of "what" is to be grasped. With such guidance, the mouth will find its way around the apple.

In this scenario, no mechanism for binding object-type and object-arrangement information needs to exist within the brain. The ventral stream provides the identification needed to select motor actions ("look around" vs. "maintain fixation" and "grasp whatever is fixated"), while the dorsal provides the sensory information needed to guide these actions. The two streams, working together through a spatial fixation strategy, can accomplish the goal of the behavior.

Animals can, of course, direct actions at objects which are not fixated. Monkeys can be trained to look at one location while attending, and possibly interacting with, another (Reynolds, Chelazzi, & Desimone, 1999; Colby & Goldberg, 1999). Does this ability imply a mechanism for binding of internal representations? Again, perhaps not. The strategy of binding through fixation may be extended toward the non-fixated case using the "searchlight" of selective attention. Within the traditional "information-processing" view of brain function, the shifting focus of attention has been seen as a mechanism for dealing with the limited capacity of the perceptual systems. Neumann (1990) argues that this has things backwards. Selection is not needed because capacity is limited. Instead, capacity is limited because selectivity is required to guide action; it is a deliberate feature of brain function which serves useful roles. One of these roles is the action-oriented filtering of stimulus information, i.e., one should process only that information which is going to be relevant to the task at hand (thus helping to make behavioral decisions already within the perceptual system). Another useful role for selectivity is binding.

Selective attention, at least in vision, is inherently spatial (see Neumann, 1990 for review). This means that it, like the fovea, may be used to bring both ventral and dorsal streams to process stimulus information pertaining to the same object. As in the fixated case, object identification in the ventral stream can direct the action of the dorsal stream upon the object of attention (by enhancing dorsal activity corresponding to the attended spatial location), and no sharing of information is necessary to grasp the apple<1>. However, such "binding through attention" requires that the relevant visual areas are in register with each other, all attending to the same spatial region. Two cases of parietal damage illustrate what can go wrong: One patient cannot shift the focus of her attention away from the fovea, and is unable to point to any location other than the one she is fixating (Carey, Coleman, & Della Sala, 1997). Such "magnetic misreaching" demonstrates the attraction that the fixation point holds for the hand. Another parietal patient exhibits great difficulty in binding stimulus features together, miscombining colors and shapes (Friedman-Hill, Robertson, & Treisman, 1995). The fact that his deficit is much worse when objects are simultaneously presented side-by-side, rather than sequentially, implicates the use of spatial information for binding stimulus-related information, even in tasks presumably performed by the ventral stream.

To summarize, certain simple behavioral tasks such as goal-directed reaching movements may be accomplished without the use of a dedicated dynamic binding mechanism. The dorsal and ventral systems can work together by focusing on stimulus information

pertaining to the same external object through a strategy of spatial selection. The ventral stream can specialize in identification of objects toward the goal of decisions, and pass these decisions to the dorsal stream simply as commands to "grasp the attended object" without needing to provide information on that object's identity. Such a strategy is robust because it exploits one of the most reliable properties of the natural world: that different objects cannot occupy the same place at the same time.

4. Action Specification and Action Selection

The above discussion points out that one of the ultimate uses of ventral information is for helping to decide among alternative courses of action. The properties of a fixated or attended object (edible vs. non-edible) may determine the kind of action which might be performed (approach vs. look elsewhere). Therefore, *both* ventral and dorsal visual systems can contribute to overt behavior.

A behaving animal has to address, at any given moment, two general kinds of questions: "what to do?" and "how to do it?". Because effectors are limited (you only have two hands, you can only run in one direction at once, etc.) an animal can perform only a small subset of all the actions possible at any given time. That is, a *selection* among actions has to be made, and sensory information (notably including the identity of objects) can be used toward that end. At the same time, the performance of any action requires the *specification* of that action's parameters (arm joint angles, length of stride, etc.). Sensory information can also be used toward that end. In fact, the ventral "what" visual system can be seen as contributing to the question of "what" action to perform, while the dorsal "how" visual system can contribute to the specification of "how" to perform it (Figure 1).

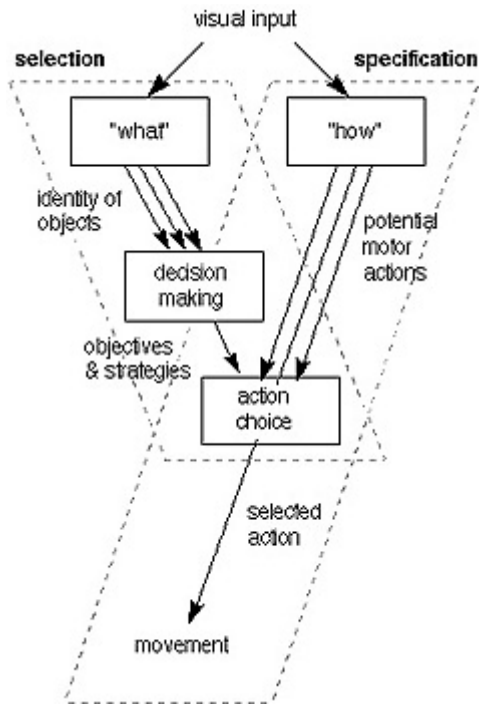


Figure 1

Specification-selection architecture for sensorimotor control (Kalaska, Sergio, & Cisek, 1998).

Both selection and specification processes can also be guided by sources of information other than the immediately available sensory input. For example, Patient D.F. (Goodale, Milner, & Carey, 1991) demonstrates that an intact ventral stream is not required for some aspects of action selection. Her ability to insert a card into a slot indicates the presence of other mechanisms (besides those related to object identification) capable of making decisions which ultimately lead to the selection of the actions performed by her visuomotor system. That is, she correctly chooses the action of inserting the card over other possible actions, such as throwing the card to the floor^{<2>}. Action specification also involves information other than current sensory input, as demonstrated by the influence of practice and familiarity upon the precision of kinetic control - this is usually discussed in terms of "internal models" of the motor plant and/or the manipulated object (Miall & Wolpert, 1996; Shadmehr & Mussa-Ivaldi, 1994).

The distinction between information useful for action selection and information for action specification does not, of course, only apply to the visual system. Information from a great variety of sources may be used to select among actions - it can include many sensory modalities as well as internal states like physiological needs and memories^{<3>}. Likewise, action specification can utilize a variety of information sources. Below, we focus on audition, and present some examples of how auditory information can be used not only to identify sound sources, but also to help guide motor performance.

5. Audition in Action

In primates, vision only collects information from the part of space toward which gaze is directed. Primate audition, in contrast, is for the most part orientation-independent. For that reason, among others, audition is extremely useful for initial orienting responses. Consider the case of hearing a loud sound. The perceived location of the sound specifies a head movement for bringing the sound source into the visual field. At the same time, other features of the sound (e.g. its amplitude) select orientation as the action most pertinent at the current time. Once the orienting movement is performed, vision can be used to select and specify subsequent actions, based upon visual properties of the presumed source. If the sound is very close or very threatening, it may specify and select an action of immediate escape away from the source. In these examples, audition is used for *both* action selection and action specification.

Beyond the triggering of an orientation response, the acoustic signal can be used to dynamically guide an ongoing action. Consider what happens after a gazelle spots a pursuing cheetah and begins to flee. Because there is no time to look back, the gazelle must rely mostly on auditory information to adjust the direction and speed of its flight. For instance, if the rhythm of pursuit of the cheetah accelerates, then the gazelle must accelerate *as much as* its predator. The bursts of energy expended by the gazelle must be timed with respect to the bursts of energy of the cheetah. Apart from the rhythm of the steps, their relative loudness provides to the gazelle some information about the distance of the cheetah. Changes in the relative timing and loudness of sequential sound events like individual steps are *spectro-temporal* properties of sounds, that is, they involve changes over time in their *spectrum* (an intensity-by-frequency description of sounds).

This excursion into the african savannah shows that the auditory signal can be used for the ongoing adjustment of the speed and/or direction of movement in a flight. In this example, vision is used for action selection (i.e., to select flight as the appropriate action) and audition, for action specification (i.e., to specify the parameters of the flight action, like the relative timing of the steps). However, unlike the initiation of an orientation response, which is an isolated motor event in time, such an example of audiomotor control involves a sequence of motor events in time, namely, the rhythmic behavior of running steps.

Conversely, in some cases audition can be used for action selection and vision for action specification. Consider the situation in which the sound of a fish diving in the water initiates the visual tracking of its swimming trajectory by a kingfisher nearby. In this example, the kingfisher selects the action of tracking based upon a startling sound, and follows the spatio-temporal trajectory of its target based upon vision. Such ecological considerations suggest that both vision and audition play a role in both action selection and action specification. In specifying the parameters of an action, vision and audition might be complementary, *vision being more informative about spatio-temporal change while audition is more informative for spectro-temporal change*. This is not to imply that vision does not contribute information on spectro-temporal change or that audition does

not contribute information on spatio-temporal change, but merely that the physical properties of visual and auditory signals motivate some degree of specialization.

In addition, there is no counterpart to the exhaustive literature on primate visuomotor control. The research on audiomotor control is mostly restricted to recent studies on animal physiology (Endepols & Walkowiak, 1999; Herbert, Klepper, & Otswald, 1997; Luksch & Walkowiak, 1998). However, we have all experienced the ease with which an auditory signal can influence the temporal parameters of our actions. Suppose that we're waiting at a red light and the car next to ours is playing music. In such situations, our hands quite easily synchronize their tapping with the rhythm of that music. In contrast, how often do we synchronize our tapping hands to the rhythm of a blinking turn signal of the car ahead of ours?

Given such considerations, one might suggest that the main contribution of audition to action specification (sounds for the "how") might be to *improve the precision with which the temporal parameters of actions are specified*. An experiment to test this hypothesis is illustrated in Figure 2, where subjects are asked to reproduce a rhythm by pressing a key in synchrony with the rhythmic signal. In half of the cases, the target rhythm is specified by a flashing light and in the other half, by a percussive sound. The visual target rhythm is presented alone, in the presence of a temporally correlated auditory signal, or in the presence of an uncorrelated auditory signal. Likewise, in the other cases, the auditory target rhythm is alone, or along with a correlated or uncorrelated visual signal. Performance is measured with respect to a baseline established by performance of the target rhythm from memory. The prediction, schematized in Figure 2, is that the addition of the visual signal will not significantly improve or disrupt the reproduction of a rhythm presented through audition, while the addition of the auditory signal will have a major effect on the reproduction of a rhythm presented through vision. In other words, the prediction is that audition will dominate temporal performance. Such an experiment is in preparation.

a) Target Rhythm	Task = Reproduction of Visual Rhythm			
 1 2 3 Time (s)	No signal (reproduction by memory)	Target visual signal only	Visual plus correlated auditory	Visual plus uncorrelated auditory
Predicted performance relative to baseline:	4 5 6 Baseline performance	4 5 6 Better than baseline	4 5 6 Much better than baseline (strong facilitation)	4 5 6 Close to baseline (strong interference)
b) Target Rhythm	Task = Reproduction of Auditory Rhythm			
 1 2 3 Time (s)	No signal (reproduction by memory)	Target auditory signal only	Auditory plus correlated visual	Auditory plus uncorrelated visual
Predicted performance relative to baseline:	4 5 6 Baseline performance	4 5 6 Much better than baseline (near perfect)	4 5 6 Much better than baseline (visual signal provides little or no facilitation)	4 5 6 Better than baseline (visual signal produces weak interference)

Figure 2

Experimental task that could be used to measure the effect of an auditory or a visual signal on the accuracy of reproduction of a rhythm presented either in the same modality or in the other one (vision and audition for the "how"). Visual signals are depicted as open circles, auditory signals as filled circles. The reproduction of the visual rhythm (a) or the auditory rhythm (b) starts after the target rhythm is presented. The accuracy of rhythm reproduction is evaluated either from memory or in the presence of the target rhythm. Consequently, any change in performance relative to the baseline results from the presence of the visual and/or auditory signal, rather than reflecting differences in memory for sequential events. In the case of reproducing a rhythm in the presence of an uncorrelated signal in the other modality (rightmost panels), there should thus be more interference from the auditory signal in the reproduction of a visual rhythm than the reverse.

6. Audition in Perception

Of course, apart from regulating the motor control of a sequential behavior such as the rhythm of locomotion (sounds for the "how"), the acoustic signal is used for sound-source identification (sounds for the "what"), such as identifying a dog from its bark or a horse from its galloping steps. It is also used for the perceptual grouping of sounds coming from a single environmental source and the segregation of sounds coming from different sources (Bregman, 1990). This is a kind of "within-system" binding problem, whose solution may also benefit by exploiting consistent properties of the environment.

Although it is likely for sounds that are produced by different environmental sources to have some degree of temporal overlap, it is unlikely that they happen to start and stop at *exactly the same time*. This regularity is so robust that sound components with a simultaneous onset are perceived as the same sound event even when other properties such as the spatial separation of their sources and/or their frequency separation suggest to the system the presence of more than one sound source (Perrott, 1984; Turgeon, 1996; Turgeon & Bregman, 1997). Conversely, the detection of an onset asynchrony between simultaneous sound components promotes their perceptual segregation into separate sound events (Darwin & Carlyon, 1995), whether they come from the same spatial location or not (Turgeon, 1996; Turgeon & Bregman, 1997). Using onset asynchrony as a segregating cue is highly adaptive in audition since it reliably signals to the system the presence of more than one environmental source.

Most of perceptual research has focused on the perception of some form of spatial change; whether it involves some change of position over time (motion), some spatial change in the absolute amount of reflected light (luminance), its differential amount (contrast) or its wavelengths (perceived color). The above excursions into the auditory domain emphasize the importance of considering spectro-temporal change in addition to spatio-temporal change in the perception of separate world objects and events. They also suggest the intriguing possibility that there may exist selective impairments in perception and action in modalities other than vision [4](#), for example, an auditory counterpart to patient D. F. (Goodale, Milner, & Carey, 1991). The literature on auditory agnosia (sounds for the "what") offers no counterpart to cases of selective impairment in identifying and recognizing objects with preserved visuomotor control. The literature on "auditory agnosia" is rather messy; the term is used for cases of phonemic decoding disorders (Korkman, Granstrom, Appelqvist, & Liukkonen, 1998), cases of selective impairments in the perception and recognition of environmental sounds (Lambert, Eustache, Lechevalier, Rossa, & Viader, 1989) and problems with the identification of musical sounds, such as melodies (Peretz, Kolinsky, Tramo, et al., 1994). Recently, there has been a reported case of a selective deficit in the perception of apparent-source movement and in the perception of rapid temporal sequences (Griffiths, Rees, Witton, Cross, Shakir, & Green, 1997). These perceptual abilities would play an important role in audiomotor control, or sounds for the "how" (e.g. the above example of the gazelle's audio-guided locomotion). Conversely, the reported cases of selective deficits in the perception and recognition of environmental sounds are consistent with problems in the use of sounds for the "what". There might exist cases of selective impairment in audiomotor control with a preserved ability to identify sound sources; conversely there might be cases of severe auditory agnosia with intact audiomotor control. There is no clinical report of such a dissociation in the neuropsychology of audition literature (personal communication with Robert Zatorre and Virginia Penhune, Montreal Neurological Institute, Montreal), but to our knowledge, this possibility has not been deliberately tested for.

7. The Specialization and Cooperation of Audition and Vision for Perception and Action

Because of the different physical properties of the visual and auditory signal (Julesz & Hirsh, 1972), vision and audition might be best suited to describe different perceptual properties of world objects and events ("what") and to specify different parameters of action ("how"): vision being specialized for spatio-temporal change and audition, for spectro-temporal change. This is not to imply that their contributions are exclusive. Just as the acoustic signal provides some information to the animal about the displacement of a target sound source in space, the visual signal can provide information about spectro-temporal change, such as a change in color over time or a change in the flickering frequency of a light. However, spatial resolution is much better in vision than in audition and temporal resolution much better in audition than vision (Perrott, 1984). There is also empirical support that spectro-temporal regularities are weighted more in auditory scene analysis than the spatial positions of sound sources: synchrony playing the dominant role in grouping sounds (Turgeon, 1996; Turgeon & Bregman, 1997).

The functional specialization of audition and vision for the spectro-temporal and spatio-temporal parameters of perception ("what") and action ("how") implies their necessary cooperation, just as in the case of the ventral and dorsal visual pathways. The auditory and somatosensory primary cortical areas project to many areas of the temporal and parietal lobes (ventral and dorsal pathways) (Pandya & Yeterian, 1985); there are also many polymodal areas in the superior and inferior temporal lobe ("what" pathway) which are connected with auditory, somatosensory and visual areas (Galaburda & Pandya, 1983). The anatomy of the mammalian brain is thus compatible with a functional distinction between a ventral multimodal perception pathway and a dorsal, sensorimotor pathway.

8. More to the "What" and the "How" than Meets the Eye

A major contribution of Milner and Goodale's book is that it is part of a growing movement away from representation-centered brain theory and toward action-centered brain theory. The importance of this movement cannot be overemphasized. Perceptual science has traditionally defined the purpose of the senses, and of vision in particular, as the construction of an internal representation of the external world. Such a representation had been assumed to be a necessary prerequisite to action control. The assumption of a unified representation, strengthened by our subjective phenomenal experience, has led to the conceptual isolation of issues of perception from issues of action. Such isolation prevents one from considering properties of the environment, and of an agent's interaction with that environment, which make solutions to certain practical problems much simpler (The example of "binding through the fovea" illustrates one such potential

solution, though we do not suggest that all kinds of binding problems can be resolved through similar strategies). And practical problems, such as grasping apples, are of more immediate relevance to the evolutionary process which created the brain than the putative problems of unified internal representations.

Just as one should not ignore the contribution of environmental regularities toward adaptive behavior, so one should not neglect the contribution of modalities other than vision. The tasks of controlling behavior, of *specifying the parameters of actions* and of *selecting among actions*, benefit by exploiting all the sensory signals available to the organism and all the reliable properties of the physical media which carry these signals. The senses may all contribute, in different degrees, to answering the questions of "what to do", "when to do it", and "how to do it".

Notes

<1> In the simplest scenario, the ventral stream can be completely insensitive to the spatial locations of objects, and merely issue commands such as "grasp what is fixated/attended" or "look elsewhere". In this case, scanning the external world might proceed completely at random, with random images falling upon the retina until something of interest is seen. However, one expects that advanced animals such as primates can do better. If the ventral system has some rough spatial sensitivity, it can bias the dorsal system for gaze orientation to preferentially select interesting objects to look at. There is a great deal of evidence that in the lateral intraparietal area (LIP), spatial information is biased by object salience (Colby & Goldberg, 1999), indicating some ventral influences upon dorsal processing without overt orientation movements.

<2> Conversely, while she is able to grasp common tools, her ventral impairment causes her to use grip arrangements which do not conform to the usual way that the tools are grasped when they are used (Carey, Harvey, & Milner, 1996). She selects one particular grasp from among the possibilities, but her selection of the grasp point is not appropriate for using the tool. That is, she selects an action of grasping in general, but she is unable to use visual information on object identity to appropriately select a specific grasp. After extensive tactile exploration, however, she is able to identify the object and correctly demonstrate its use.

<3> What about conscious perception? Milner and Goodale (1995) suggest that only ventral processing reaches conscious awareness. Bernard Baars (1988) has suggested that conscious awareness is almost exclusively perceptual. In our framework, these suggestions translate to the proposal that among the action selection and action specification systems, it is only the perceptual information ultimately leading to action selection which enters conscious awareness, and not that which leads to the specification of action parameters. Why might this be so? One potential answer comes from a consideration of what consciousness might be for, in the first place. Though the debate on this issue will surely continue for many years, one line of speculation suggests that

consciousness provides an internal narrative which serves to describe one's behavior in terms of a story about purpose (Dennett, 1991). If that is the case, then one might suggest that such a story is useful for organizing action selection, preventing current actions from undoing what past actions had accomplished (Cisek, 1997). For that reason, only the perceptual cues used to select among actions have to be woven into the conscious narrative, while the details of action specification may be left out.

<4> Indeed, such a dissociation has been shown for somatosensation. Paillard, Michel, and Stelmach (1983) describe a patient who is unable to consciously detect touch on her deafferented hand even though she is capable of reaching to the point of contact.

Acknowledgements

The authors thank David Carey and an anonymous reviewer for pointers to relevant literature and helpful suggestions which have improved this commentary. P. C. supported by NIH, M. T. supported by NSERC.

References

- Baars, B.J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Grouping of Sounds*. Cambridge, MA: MIT Press.
- Carey, D.P. (1997). Action, perception, cognition, and the inferior parietal cortex. *Trends in Cognitive Sciences*, 2, 162-164.
- Carey, D.P., Coleman, R.J., & Della Sala, S. (1997). Magnetic misreaching. *Cortex*, 33, 639-652.
- 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.
- Cisek, P. (1997). Global Workspace theory in the spotlight of evolution. *Journal of Consciousness Studies*, 4, 310-313.
- Colby, C.L., & Goldberg, M.E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319-349.
- Darwin, C.J., & Carlyon, R.P. (1995). Auditory grouping. In B.C.J. Moore (Ed.), *Hearing: Handbook of Perception and Cognition*. (pp. 387-424). London: Academic.

- Dennett, D.C. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.
- Endepols, H., & Walkowiak, W. (1999). Influence of descending forebrain projections on processing of acoustic signals and audiomotor integration in the anuran midbrain. *European Journal of Morphology*, *37*, 182-184.
- Friedman-Hill, S.R., Robertson, L.C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science*, *269*, 853-855.
- Galaburda, A.M., & Pandya, D.N. (1983). The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkeys. *Journal of Comparative Neurology*, *221*, 169-184.
- Goodale, M.A. (1996). Visuomotor modules in the vertebrate brain. *Canadian Journal of Physiology and Pharmacology*, *74*, 390-400.
- 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., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154-156.
- Griffiths, T.D., Rees, A., Witton, C., Cross, P.M., Shakir, R.A., & Green, G.G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction: A psychophysical study. *Brain*, *120*, 785-794.
- Herbert, H., Klepper, A., & Otswald, J. (1997). Afferent and efferent connections of the ventrolateral tegmental area in the rat. *Anatomy and Embryology*, *196*, 235-259.
- Joliot, M., Ribary, U., & Llinas, R. (1994). Neuromagnetic coherent oscillatory activity in the vicinity of 40-Hz coexists with cognitive temporal binding in the human brain. *Proceedings of the National Academy of Sciences USA*, *91*, 11748-11751.
- Julesz, B., & Hirsh, I.J. (1972). Visual and auditory perception - An essay of comparison. In E.E. David, Jr. & P.B. Denes (Eds.), *Human Communication: A Unified View*. (pp. 283-340). New York: McGraw-Hill.
- Kalaska, J.F., Sergio, L.E., & Cisek, P. (1998). Cortical control of whole-arm motor tasks. In M. Glickstein (Ed.), *Sensory Guidance of Movement, Novartis Foundation Symposium #218*. (pp. 176-201). Chichester, UK: John Wiley & Sons.
- Korkman, M., Granstrom, M.L., Appelqvist, K., & Liukkonen, E. (1998). Neuropsychological characteristics of five children with the Laudau-Kleffner syndrome: dissociation of auditory and phonological discrimination. *Journal of the International Neuropsychological Society*, *4*, 566-575.

Lambert, J., Eustache, F., Lechevalier, B., Rossa, Y., & Viader, F. (1989). Auditory agnosia with relative sparing of speech perception. *Cortex*, 25, 71-92.

Llinas, R., Ribary, U., Joliot, M., & Wang, X.-J. (1994). Content and context in temporal thalamocortical binding. In G. Buzaki, R. Llinas, W. Singer, A. Berthoz, & Y. Christen (Eds.), *Temporal Coding in the Brain*. (pp. 251-272). Berlin, Heidelberg: Springer-Verlag.

Luksch, H., & Walkowiak, W. (1998). Morphology and axonal projection patterns of auditory neurons in the midbrain of the painted frog, *Discoglossus pictus*. *Hearing Research*, 122, 1-17.

Marr, D.C. (1982). *Vision*. San Francisco: W. H. Freeman.

Mattingley, J.B. (1999). Attention, consciousness, and the damaged brain: Insights from parietal neglect and extinction. *Psyche*, 5. <http://psyche.cs.monash.edu.au/v5/psyche-5-14-mattingley.html>.

Miall, R.C., & Wolpert, D.M. (1996). Forward models for physiological motor control. *Neural Networks*, 9, 1265-1279.

Milner, A.D., & Goodale, M.A. (1995). *The Visual Brain in Action*. Oxford University Press.

Milner, P.M. (1974). A model for visual shape recognition. *Psychological Review*, 81, 521-535.

Neumann, O. (1990). Visual attention and action. In O. Neumann & W. Prinz (Eds.), *Relationships Between Perception and Action: Current Approaches*. (pp. 227-267). Berlin: Springer-Verlag.

Paillard, J., Michel, F., & Stelmach, G. (1983). Localization without content: A tactile analogue of 'blind sight'. *Archives of Neurology*, 40, 548-551.

Pandya, D.N., & Yeterian, E.H. (1985). Architecture and connections of cortical association areas. In A. Peters & E.G. Jones (Eds.), *Cerebral Cortex*. (pp. 3-61). New York: Plenum.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117, 1283-1301.

Perrott, D. (1984). Concurrent minimum angle: a re-examination of the concept of auditory spatial acuity. *Journal of the Acoustical Society of America*, 75, 1201-1206.

Reynolds, J.H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736-1753.

Roelfsema, P.R., Engel, A.K., Konig, P., & Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, *385*, 157-161.

Shadmehr, R., & Mussa-Ivaldi, F.A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, *14*, 3208-3224.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings. *Behavioral and Brain Sciences*, *16*, 417-494.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In R. Llinas & P.S. Churchland (Eds.), *The Mind-Brain Continuum: Sensory Processes*. (pp. 101-130). Cambridge, MA: MIT Press.

Turgeon, M. (1996). Rhythmic Masking Release: A paradigm to investigate the auditory organization of tonal sequences. In *Proceedings of the 4th International Conference on Music Perception and Cognition*. (pp.315-316). Montreal: McGill University.

Turgeon, M., & Bregman, A.S. (1997). 'Rhythmic Masking Release': A paradigm to investigate auditory grouping resulting from the integration of time-varying intensity levels across frequency and across ears. *Journal of the Acoustical Society of America (Suppl.)*, *102*, 3160

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. 549-586). Cambridge, MA: MIT Press.

von der Malsburg, C. (1996). The binding problem of neural networks. In R. Llinas & P.S. Churchland (Eds.), *The Mind-Brain Continuum: Sensory Processes*. (pp. 131-146). Cambridge, MA: MIT Press.