

"Visual Awareness and Visuomotor Action" JOURNAL OF CONSCIOUSNESS STUDIES 6:11-12: 1999.  
p.1-18

# **Visual Awareness and Visuomotor Action**

**Andy Clark**  
**Philosophy/Neuroscience/Psychology**  
**(PNP) Program**  
**Washington University in St. Louis**  
**USA**  
**[andy@twinearth.wustl.edu](mailto:andy@twinearth.wustl.edu)**

Thanks to: Susan Hurley, Brian Cantwell Smith, David Chalmers, Jesse Prinz, Josefa Toribio, an anonymous referee for J.C.S and all the seminar participants at the Washington University School of Medicine, and at the 1<sup>st</sup> Gulbenkian Symposium on Cognitive Neuroscience, Lisbon, July 1998.

## **ABSTRACT**

Recent work in "embodied, embedded" cognitive science links mental contents to large-scale distributed effects: dynamic patterns implicating elements of (what are traditionally seen as) sensing, reasoning and acting. Central to this approach is an idea of biological cognition as profoundly "action-oriented" - geared not to the creation of rich, passive inner models of the world, but to the cheap and efficient production of real-world action in real-world context. A case in point is Hurley's (1998) account of the profound role of motor output in fixing the contents of conscious visual awareness - an account that also emphasizes distributed vehicles and long-range dynamical loops. Such stories can seem dramatically opposed to accounts, such as Milner and Goodale (1995), that stress relatively local mechanisms and that posit firm divisions between processes of visual awareness and of visuomotor action. But such accounts, I argue, can be deeply complimentary and together illustrate an important lesson. The lesson is that cognition may be embodied and action-oriented in two distinct - but complimentary - ways. There is a way of being embodied and action-oriented that implies being closely geared to the fine-grained control of low level effectors (hands, arms, legs and so on). And there is a way of being embodied and action-oriented that implies being closely geared to gross motor intentions, current goals, and schematic motor plans. Human cognition, I suggest, is embodied and action-oriented in both these ways. But the neural systems involved, and the size and scope of the key dynamic loops, may be quite different in each case.

## **1. Local versus Highly Interactive Explanations.**

Science in general, and cognitive science in particular, seems currently torn between two superficially competing kinds of explanation.<sup>1</sup> The struggle is between explanations that highlight distributed complexity and large-scale, non-linear, interactive effects and those that highlight smaller circuits; between - to invoke an old but persistent dichotomy - holism and reductionism. In the next section, I'll show how this general tension plays out in the specific arena of (accounts of) visual awareness. But it is useful, I think, to begin with a rough sketch of the larger issues.

Thus consider the most general claim of what has come to be known as "embodied, embedded cognitive science".<sup>2</sup> The claim is that the essence of adaptive intelligence often lies in complex, non-linear processes that span multiple inner systems and include, as active contributors, aspects of body, action, and world. A frequent corollary of this claim is that there are no neat dividing lines between perception, cognition and action; that much of what looks like action plays fundamental cognitive and computational roles; and that perceptual processing phases gradually into cognition and abstract reason. To get an explanatory grip on such world - and action - exploiting systems, it has been argued<sup>3</sup>, we must attend heavily to extended feedback and feedforward loops whose (often continuous) activity underpins adaptive intelligence.

Such an expanded focus does indeed seem helpful when confronting several aspects of biological intelligence. I shall give one example, and merely gesture at several others. The example concerns the production of rhythmic motor actions, such as walking, rowing, certain kinds of coordinated finger motion, and even human speech. There is now a large and compelling literature<sup>4</sup> that shows how such activity arises from the complex interactions of neural resources, bodily bio-mechanics and (sometimes) external environmental structure. Such accounts reject the once-popular<sup>5</sup> view of "centralized control" or "central pattern-generation" which "assumes a central representation of the movement, including its form, amplitude and temporal characteristics, that is imposed on the periphery" (Hatsopoulos and Warren (1996) p.

3). Instead of seeing rhythmic motor action as the muscular expression of an inner rhythmic command, theorists such as Kelso (1995), Thelen and Smith (1994), Turvey and Kugler (1987) and, more impressionistically, Bernstein (1967), see such actions as the tuned product of neural - bodily interactions. Such accounts need not - and often do not - deny the existence of something *like* a central pattern generator. But they deny that such a resource *determines* the rhythmic notions of the embodied agent:

Any central timing process, such as a neural oscillator, is not acting as an extrinsic timekeeper to drive the peripheral segments but must be reciprocally modulated by information about the dynamics of the periphery. Preferred motor timing emerges from the interaction of central and peripheral components, rather than from being dictated by either.

Hatsopoulos and Warren (1996) p. 10

The basic mechanism proposed by Hatsopoulos and others involves the use of proprioceptive feedback (from a musculoskeletal system with its own intrinsic, spring-like dynamics) to tune the neural resources to the resonant frequency determined by the peripheral biomechanics. This tuning, which can be experimentally demonstrated<sup>6</sup>, allows the system to cope fluently with physical changes (limb growth, muscle growth, etc.) and to find "least-energy" solutions to the problem of generating rhythmic actions. Moreover, the very same mechanisms that allow a biological agent to thus couple its neural resources to properties of the bodily periphery also allow for coupling to the wider environment. Experiments in which a subject performs a task using a hand-held pendulum (Kugler and Turvey (1987)) show a selected frequency of oscillation that becomes matched to the resonant frequency of the whole wrist-pendulum system. The same effect will occur in a good golfer or racket sport player, or when rocking a car to get it out of the snow.<sup>7</sup> In all these cases, the proprioceptive information couples the neural system to bodily and/or environmental resources in a way that creates "a larger autonomous dynamical system" (Hatsopoulos and Warren (op cit) p. 12). This effect is even reflected, it now seems, in the receptive field size of cells involved in somatosensory and visual

processing. Iwamura (1998) cites Iriki et al. (1996) as showing that after a monkey repeatedly used a tool (a rake) for the retrieval of food "the visual RF<sup>s</sup> [receptive field] became elongated along the axis of the tool, as if the image of the tool was incorporated into that of the hand" (Iwamura (1998) p. 525).

Other examples of an embodied, environmentally embedded perspective include Beer's (1995) work on leg control in a simple robot 'insect', Webb's (1994) work on phonotaxis (sound-dependent tracking) in both real and robot crickets, and Thelen and Smith's (1994) work on reaching and stepping in human infants. Chiel and Beer (1997) offer a useful review of the literature on the importance of brain/body interactions, concluding that "adaptive behavior is the result of the continuous interaction between the nervous system, the body and the environment" and that "one cannot assign credit for adaptive behavior to any one piece of this coupled system" (op cit, p. 555).

It can often seem, however, as if the embodied approach really pays dividends only for relatively low-level, motoric aspects of adaptive behavior. And without a doubt, the most radical versions of the embodied approach do face special challenges as we ascend to the levels of reasoned thought, imagination and off-line planning and rehearsal (for a critical but sympathetic account, see Clark (1999)). It seems clear, however, that the broad idea of neural systems becoming deeply geared to the exploitable presence of bodily or external structure is highly applicable to certain aspects of 'advanced reason' - we can see the biological brain as coupled, via the body, to all kinds of technological and artifactual resources (pens, paper, PCs), such that (to paraphrase Chiel and Beer) "one cannot assign credit for intellectual success to any one piece of this complex system".<sup>9</sup>

In many cases, however, there is only a surface opposition between such a distributed, interactive perspective and more traditional (localist, internalist) concerns. There may, for example, be a perfectly good localist story to tell about how certain aspects of individual brain function provide the pattern-completing substrate necessary for the larger person-plus-pen-and-paper system to then function as an integrated, extended numerical computation device.

To set up a genuine *opposition* between the two visions, certain conditions need to be met. First, we need to be convinced that we really confront two different stories about the *same phenomenon* (a condition I'll call "same target"). Second, we need to be convinced that the interactive story, if it is to be preferred, is also in some important but hard-to-pin-down sense *deep* (a condition I'll call "deep embodiment").

In the rhythmic motion case just described, these conditions seem to be met. The target phenomenon was rhythmic motion generation. One (localist, internalist) story explained this as the effect of a central pattern generator whose own frequency of oscillation determined a temporally matched oscillation in, say, a limb. The other (interactionist) story denied the presence of such a closely matched inner resource, and explained the very same target motion via a complex process of reciprocal modulation in which the inner resources sought a kind of energy efficient compromise with the intrinsic dynamics of the limb and musculo-skeletal system. The "embodied" account is here a genuine "same target" competitor. But is it deep? By this I mean, does the stress on interaction illuminate the phenomenon in a truly revealing way? Consider a scenario in which interaction might matter yet (intuitively) not in a deep way. Imagine a central pattern generator with a small number of pre-determined settings (frequencies of oscillation), and suppose, in addition, that certain instances of proprioceptive feedback determines which of these (three or four) frequencies is generated. In such a case, interaction and feedback *matters*. But it matters only as a kind of input to a central system whose operation still fixes the target behavior. We might thus respond by, in essence, re-parsing the original target into two semi-independent components: a peripheral, dynamic loop, and a central system which, once informed, does the real work. Contrast this with the actual story, in which the target behavior was revealed as a genuine product of complexly interacting dynamics and in which the detailed biomechanics of the periphery continuously and sensitively influences the central resource so as to yield a signal whose effect - in biomechanical context - is to match the preferred frequency of motion to the resonant frequency of whatever peripheral system (arm, arm-plus-golf club, etc.) is in play. When the interactions are this important, continuous, and

subtle, there is - or so I claim - a very real sense in which no re-parsing can yield an internalist story which still *keeps sight* of the target phenomenon. There is, of course, an inner story to tell. But the inner story now fails to account for the subtlety, power and efficiency of the target phenomenon - a task that necessitates a more extended and interactive perspective.

Hopefully, then, we now have at least a rough sense of the kind of case in which an embodied, embedded story is a *genuine* and *deep* competitor to some localist/internalist alternative. On, then, to our target case: perceptual (and especially visual) awareness.

## **2. Two Takes on Visual Awareness: The Ventral Stream versus Escher Spaghetti**

The recent debate concerning visual awareness looks superficially similar to the one concerning rhythmic motion. Here, too, we find both localist accounts and more 'embodied' alternatives that stress larger-scale dynamic loops. A compelling example of a recognizably localist account is Milner and Goodale's (1995)(1998) account of the ventral stream correlates (see below) of visual awareness. A good example of a dynamic, interactionist account is Susan Hurley's (1998) depiction of conscious visual experience as dependent on a complex web of relations between perception and action. Brief sketches seem in order.

Milner and Goodale (1995) suggest that visual awareness depends on the activity of selected parts of the visual processing system. By "visual awareness" they mean the capacity to know and recognize objects by means of conscious, visual experience: a capacity that is taken to imply - in a normally-functioning language user - the ability to report that a visually presented object is a such-and-such, that is spatially oriented thus-and-so, and so on.

The specific claim is that these capacities of visual awareness (which they sometimes call capacities of visual *perception* and contrast with capacities for visually-guided *action*) depend on a specific visual processing stream - the ventral stream - that is said to operate semi-independently of the processing stream (the dorsal stream) that guides fine-tuned motor action in the here-and-now. The ventral stream structures that are thus claimed to support awareness include areas (V2, V3, V4, TEO) projecting to the inferior temporal cortex (IT), while the dorsal

stream story implicates areas projecting to the posterior parietal (PP) cortex (see fig. 1 attached).

Regarding these two streams the claim is that:

The visuomotor modules in the primate parietal lobe function quite independently from the occipitotemporal mechanisms generating perception-based knowledge [visual awareness] of the world.

Milner and Goodale (1998) section 4

Evidence for this deep dissociation comes in three main varieties: deficit data concerning patients with damage to areas in either the dorsal or ventral streams; performance data from normal human subjects; and computational conjectures concerning the inability of a single encoding to efficiently support both visual form recognition and visuomotor action.

The relevant deficit data concerns two classes of patients: visual form agnosics and optic ataxics. The visual form agnostic DF (who suffered from carbon monoxide poisoning causing ventral occipital damage) is unable to judge (unable to report on, and reports no visual experience of) the shape and orientation of visually presented items. But she retains significant visuomotor skills, and is able to catch a ball or stick with correct hand orientation, to place her flattened hand through a visually presented "letterbox" and so on (for a full account, see Milner and Goodale (1995) ch. 5). Optic ataxics, by contrast, have damage to the dorsal stream and, although they are visually aware of (and can report on) shape and orientation, they are unable to fluently reach for and grasp the objects they so clearly see. As Gazzaniga (1998, p. 109) has it, "it is as though they cannot use the spatial information inherent in any visual scene."

Moving to unimpaired, normal agents, Milner and Goodale appeal to various experimental results displaying a conflict between verbal judgments, based on visual experience, and the visual knowledge that is manifest in non-verbal actions. A neat example concerns the 'Ebbinghaus' or 'Tichener Circles' illusion (see fig. 2), in which visual awareness (visual experience) delivers an illusory content which nonetheless fails to inform subsequent fine-tuned motor action. To show this, Aglioti et al (1995) set up a physical version of the illusion, using poker chips as circles. Subjects were told to pick up a specific disc if they saw the two discs as



differing in size. By choosing a specific disc, subjects showed they were susceptible to the illusion, since the discs were actually identical in size. Nonetheless, by using infra-red light emitting diodes attached to finger and thumb, the act of picking up the chosen disc was shown to be finely calibrated in advance of actually touching the disc, displaying a pre-formed precision (thumb and forefinger) grip closely tuned to the actual (non-illusory) size! The explanation, according to Milner and Goodale (1995, p. 168) is that this precision grip was calculated, using non-conscious visual information, by the dorsal stream, and only the conscious, ventral stream was 'fooled'.

The deep reason for such functional compartmentalization, Milner and Goodale conjecture, involves the very different computational demands of visuomotor guidance and object recognition. The former requires precise knowledge of spatial location and orientation, and must be constantly and egocentrically updated to reflect real-world motion and relative location. The latter requires us to identify something as *the same* thing, irrespective of motion and current spatial orientation, and demands only as much spatial sensitivity as is necessary to support conscious object recognition and reasoning. The dorsal stream is thus said to be specialized for fluent motor interaction while the ventral stream deals with enduring object properties and subserves explicit recognition and semantic recall. As a kind of corollary, the ventral stream must take over whenever the real-world object is not present-at-hand: actions in respect of imagined or recalled objects are under ventral stream control, and this is reflected in grosser kinematics of grasp and sometimes (in the case of ventrally damaged subjects) total failure to perform (see Milner and Goodale (1995) pp. 136-138).

Milner and Goodale thus offer an account of the physical correlates of visual awareness which stresses the importance of a closely circumscribed subset of neural mechanisms (V2, V3, V4, TEO and IT), and excludes the occipito-parietal structures implicated in fine-grained motor control. The upshot, provocatively expressed, is that "what we think we 'see' is not what guides our actions" (Milner and Goodale (1995) p. 177).

The Milner and Goodale account belongs squarely in the tradition<sup>10</sup> of betting on relatively local neural structures and features as the likely physical correlates of visual awareness. The general alternative, as we noted, is to bet on models that stress co-ordination and interaction.<sup>11</sup> As an example - and one which now moves us all the way into the terrain of complex interactive dynamics - consider Susan Hurley's (1998) account in which the contents of perceptual experience are linked (in a deep, 'non-instrumental' way - see below) to motor outputs, and in which there is heavy stress on "multiple channels of motor feedback, some of which go through states of the environment" (op cit p. 328).

To lay the groundwork for such an account, it is necessary to recognize an important idea, central to the literature on embodied cognition, and increasingly reflected in mainstream neuroscientific conjecture. It is the idea that we should, in many cases, resist the temptation to think in terms of a simple linear flow in which the senses deliver input which is progressively processed and refined until an output (usually a motor action) is selected, and the process repeats. Hurley calls this the Input-Output picture and argues that it should be rejected in favor of accounts which give due weight to the looping and temporally continuous nature of many processes of information flow and control. For example, there is evidence that a great deal of activity in "early" visual processing areas is modulated by activity flowing back down from "higher" areas. Experimenters have found top-down context effects on the receptive fields of cells in V1, enhanced responses of cells in V1 and V2 to locations to which monkeys are about to saccade, and so on.<sup>12</sup> Moreover, as Churchland et al (1994, p. 43) point out, "all cortical areas, from the lowest to the highest, have numerous projections to lower brain centers, including motor-relevant areas such as the striatum, superior colliculus, and cerebellum". Such back-projections allow highly processed states - decisions to move, top-level semantic information, etc. - to loop back to affect the low-level processing of current inputs. Such descending connectivity lies at the heart of Edelman's (1987) account of "reentrant processing" in which such downward pathways establish correlations between multiple cortical and sub-cortical areas. Patterns of mutual influence between distinct areas, set up by the use of re-entrant connectivity,

allow activity in one site to become correlated with activity at others, and these correlated patterns of activity (according to Edelman - see also Sporns et al (1989)) are themselves the distributed vehicles that carry information about high-level features. Most high-level cognitive capacities, on this account, depend on the correlated activity of multiple neural areas, including motor areas. Similar themes surface in the "convergence zone" hypothesis of Damasio and Damasio (1994). Convergence zones are local brain areas whose task is not to directly encode information but to enable correlated activity to occur in multiple neural sub-systems. A convergence zone is thus an area in which several long-range corticocortical connections (feedback and feedforward) converge, enabling signals from that area to simultaneously influence multiple distant regions of neural tissue. And once again, various types of high-level knowledge are said to depend on various kinds of correlated activity, orchestrated by a kind of hierarchy of convergence zones (see eg. Damasio and Damasio, op cit p. 73).

Many theorists, it thus seems, are stressing the looping dynamics of inner processing, and describing cognitive mechanisms in which "information flows back as it flows up, and it flows more or less continuously" (Hardcastle 1998, p. 341).<sup>13</sup> The physical *vehicles* of certain kinds of higher level content, on these accounts, are often extended dynamic loops connecting 'higher' to 'lower' brain areas, and encompassing both 'cognitive' and 'motor' systems.<sup>14</sup> It is an image that I think of as Escher Spaghetti<sup>15</sup> - not just multiple criss-crossing strands (ordinary spaghetti), but strands whose ends feed back into their own (and others) beginnings, making 'input' and 'output', and 'early' and 'late' into imprecise and misleading visions of complex recurrent and reentrant dynamics.

Hurley (1998) is, in large part, a sustained application of these kinds of ideas to issues concerning perceptual awareness and its relation to motor output. The key claim is that such perception can depend *non-instrumentally* on motor output. Perception depends merely *instrumentally* on output when, for example, you turn your head and see something new. The new perceptual content is (trivially) made available by your action. The cases Hurley is chasing, however, are ones in which what matters is not this pragmatic effect of the current output (action

command) on the next input, but a more profound effect of the output signal *itself* on perceived content: "When perceptual content depends *non-instrumentally* on output, it does not do so via input, but directly" (Hurley 1998, p. 342). One mark of genuine non-instrumental dependence, for Hurley, seems to be the capacity to vary the perceptual content by a change in the motor output signal *even though the current input remains fixed*. The goal is to capture "the thought that perception and action are in some way co-constituted" (op cit p. 342).

The very simplest example of the kind of thing Hurley has in mind is the case of the paralyzed eye. According to Gallistel (1980), if someone with paralyzed eye muscles tries to look (say) to the right, the eye does not respond (so there is no change in gross retinal image) but the world, as presented in visual experience, appears to jump to the right. Hurley concludes that since there is a change in perceptual (conscious) content, and no *other* apparent difference except the introduction of the motor signal, the difference in content is best explained by a non-instrumental role for the motor signal itself (op cit p. 372). Several other cases are discussed, including cases involving TVSS (tactile visual substitution systems), adaptation to left-right reversing goggles, cases of output neglect and post-commissurotomy experiments (op cit, ch. 9). In the latter case, a patient with a severed corpus callosum attempts to perform a left-hand task relating to a card that is visually presented in the right visual field, while fixating a central point. The patient reports that:

The image of the object, initially seen, was blotted out of awareness in the left hemisphere the instant a movement, initiated by the right hemisphere, had started.  
Trevarthen, 1984, p.333  
quoted in Hurley 1998, p. 355

Hurley's interpretation is that "the motor intention relating to the left hand brought sensory information from the left visual field to perceptual awareness" (Hurley 1998, p. 374), and adds that in this case "not just the content of consciousness, but also the presence or absence of consciousness, may depend on relations of input to output" (op cit. p. 374). Once again, we seem to see the "motor-dependent selection of which information is present to perceptual

consciousness" (op cit p. 365). From such cases, Hurley is led to conjecture that the perceptual content (the content available to perceptual awareness) depends on an entire "feedback relationship between input and output" (op cit p. 375).

In the case of the paralyzed eye, this feedback loop must be wholly internal, perhaps involving the influence of efferent copy (functioning as a kind of "virtual input") on the visual processing stream. But, as Hurley frequently stresses, feedback loops may also run through the external environment. This would be the case if, for example, perceptual content depended at times on "patterns of afference that are a function of movement through the environment" (op cit p. 416) - a view often stressed in ecological views of perception. Such world-involving feedback loops will count as issuing in non-instrumental dependence if (and only if) what fixes the perceptual content is not *just* the sequence of inputs dictated by the motions (as Gibson thought) but rather the *relation* between motion *signals* and resultant inputs (i.e. passive motion won't have the same effect). It is thus the putative involvement of the motor output signal *itself* (and not its effects on subsequent worldly input) in determining the contents of our perceptual experience that makes for the *non-instrumental* dependence of perception on motor output, and that allows, more generally, for a genuinely *deep* (see section 1 above) dependence of perceptual content upon "feedback loops with orbits of varying sizes, ... that can...in principle spread across internal and external boundaries" (op cit p. 327). The full conjecture is thus that perceptual contents (including the contents of visual awareness) may sometimes depend on whole loops involving motor signals, body and world. It is a picture which, as Hurley (op cit p. 20) notes, "takes the notion of distributed processing to its logical extreme": just as the physical bearer of a given content might be a pattern of activation across a whole neural population, so too it might be a pattern of relations *between* neural populations, or between sensory inputs, motor outputs, and external states of affairs.

We seem to thus confront a genuine conflict. Milner and Goodale offer a localist, internalist account of visual awareness, which draws a firm distinction between mechanisms of visual awareness and mechanisms of visuomotor action. While Hurley offers a highly

interactionist picture in which the contents of visual awareness are *deeply* bound up with motor output and (perhaps) bodily and external states of affairs.

### 3. Two Ways to be Action-Oriented

Is the conflict depicted in the previous section real or merely superficial?<sup>16</sup> Hurley's story does indeed involve what we termed (in section 1) a "deep" role for large-scale dynamics. But is it really a "same target" competitor? Hurley's principal conjecture, recall, is that the content of the visual experience would not be as it is were it not for the non-instrumental role of motor output: a conjecture echoed by, eg. Mandik (1999), who argues that the introspectible properties of visual experience (the qualia, on his definition) are:

determined not only by the nature of the information transduced by the nerve endings in the sensory organs, but also by what type of subsequent motor activity that information is employed in.

Mandik (1999) p. 56

But now consider Milner and Goodale's claim in just a little more detail. The ventral processing stream, they argue, is the physical locus of visual awareness, while the dorsal stream guides fine-grained visuomotor action in the here-and-now. Clearly, though, there must be interactions between the two.<sup>17</sup> I can choose to reach for an item because it looks (phenomenally, in my visual experience) larger than another, or redder, or whatever. The two streams anatomically exhibit multiple kinds of looping cross-connectivity, and there is ample evidence of information transfer between the two. Milner and Goodale's suggestion (1995, pp. 201-204) is that the ventral stream - the one said to be subserving visual experience - helps select both the *type* of action to perform and the *object* upon which to perform it.

Thus recall the Titchener Circles experiment. Here, according to Milner and Goodale, the choice of *which* disc to pick up, and the choice of what *kind* of action to perform (picking it up and not, e.g., poking at it), both depend on conscious, ventral stream dominated processing. It is only the most fine-grained visuomotor performance (the specifics of the precision grip) that is then deferred to full dorsal stream control. Let's call the choice of targets for action and the

choice of type of action "gross motor selection", and let's call fine-grained visuomotor control "fine motor tuning". The possibility then remains open - or so it seems to me - that acts of gross motor selection might actually *influence* the conscious content of the visual experience even if fine motor tuning remains relatively insulated. It remains possible, for example, that had the gross motor command selected different action-types or different goal-objects, the visual experience would have differed, and done so non-instrumentally, i.e. without the need for any change in gross input. In short, just because *fine motor tuning* is insulated from perceptual experience, it doesn't follow that perceptual experience is independent of *all types* of motor output signal: in particular, it may not be independent of the grosser – and ventral stream involving - events of action-type and target-object selection.

The most general lesson here is that the notion of an inner state or process being “action-oriented” is by no means as simple as it sounds. The kinds of encoding supported by the dorsal stream are, as Milner and Goodale point out, heavily geared to what we have called “fine motor tuning”. They are geared to action in the sense of being geared to precise and egocentrically defined *movements*, which must themselves be geared to the real locations, sizes and orientations of objects in space. The kinds of encoding supported (in part) by the ventral stream are geared to action in quite a different sense. They are geared to *action* in the sense in which actions reflect the specific needs, purposes and intentions of the agent. What is required here is an encoding that captures how the surrounding world is in just those respects necessary for planning, choice and conscious reasoning. Milner and Goodale's prime conjecture is that the computational demands of the latter complex of activities (planning, choice and reason) are dramatically distinct from – and even inconsistent with – those of the former (fine motor control). For the reasoning and planning complex requires us to identify objects regardless of spatial orientation and absolute location, and regardless of the current disposition of our limbs and bodies; and it demands only rough and relative information about spatial matters (what is closer, further, what is next to what, and so on). In this vein, Goodale and Humphrey recently speculate that:

“...if perceptual representations [the ones underlying conscious visual awareness] were to attempt to deliver the real metrics of all objects in the visual array, the computational load would be astronomical. The solution that perception [visual awareness] seems to have adopted is to use world-based co-ordinates – in which the real metric of that world need not be computed. Only the relative position, orientation, size and motion of objects is [here] of concern. For example, we can watch the same scene unfold on television or on a movie screen without being confused by the enormous change in the co-ordinate frame. [But] as soon as we direct a motor act towards an object, an entirely different set of constraints applies.”

Goodale and Humphreys (1998) pp. 195-196

I have quoted this passage at length since it is, I think, very revealing. Visual awareness, if this story is correct, simply cannot *afford* to be action-oriented in the first of our two senses. It simply cannot afford to represent each and every aspect of the scene present in visual awareness in the precise and egocentrically defined co-ordinates required to support complex physical interactions with that very scene. But the pressure for computational economy may very well drive the systems underlying visual awareness to *be* action-oriented in the *second* sense we described. They are action-oriented insofar as they register just those aspects of the visual scene relevant to our current needs, purposes and motor plans.

There is, indeed, suggestive evidence that conscious visual awareness is action-oriented and motor-involving in this latter sense. Hurley’s own example of the visual field effects in the post-commissurotomy patient suggest a system which (when functioning normally) ties visual awareness to the locations indicated by current motor plans. While the growing body of evidence concerning so-called change-blindness (Simons and Levin (1997)) seems to show that conscious visual awareness is not in the business of building and maintaining a detailed inner model of the visual scene. (In a typical change-blindness experiment, elements of a visually present scene are altered while the subject saccades around the display. Subjects prove remarkably insensitive to these changes, leading theorists (such as Ramachandran et al (1994) to conclude that the feeling



we have of perceiving a richly detailed scene is due to our capacity to repeatedly scan and foveate the visual scene so as to extract task-relevant information just as it is needed (but not a moment before)). For some lively demonstrations, see Ballard et al (1997).

Another factor that probably contributes to the sparse yet gross-action-oriented nature of the contents of visual awareness is the surprisingly tight relation between visual awareness and certain processes of *attention*. Thus consider the fascinating claim that "there is no conscious perception at all in the absence of attention" (Mack and Rock, 1995, p. 227). Mack and Rock are driven to this claim by multiple experimental routes, including (especially) the observation that in the presence of an attention-diverting stimulus, a stimulus which would otherwise be seen and reported is very often not consciously perceived at all: an effect which is actually *magnified* if the critical (nondiverting) stimulus is foveally presented. For example, if the critical stimulus was a small, black square presented while the viewer was asked to report the longer arm of a briefly presented cross, the square was not consciously perceived in 25% of cases in which the critical stimulus was parafoveally presented, and in a full 75% of cases where it was foveally presented! By contrast, in experiments with no diverting stimulus or in conditions of divided attention, most subjects could perceive the square. Mack and Rock dub this phenomenon "inattentional blindness" and conclude, quite generally, that it is only once attention is engaged by an object that the object becomes "visible" to conscious awareness (see e.g. Mack and Rock, op cit, pp. 227-228). This may seem impossible - how can attention be drawn to that which is not yet perceived? But it makes good sense once we distinguish nonconscious sensory pickup from conscious perceptual awareness. Indeed, it is well-known - and Mack and Rock's own subsequent experiments<sup>18</sup> also reflect this - that stimuli that do not make it into conscious awareness may nonetheless be highly processed, and this information can be used to guide behavior. Indeed, it is often necessary for a stimulus to be highly processed (yielding an implicit grasp of abstract meaning and significance for the agent) *before* attentive processes render it consciously available<sup>19</sup> (see Mack and Rock, op cit, p. 229).

The link between visual awareness and attention is, I believe<sup>20</sup>, deep and significant. It fits nicely, moreover, with the compelling idea (Baars (1998), Dennett (1991), Chalmers (1998)) that conscious awareness is bound up with processes (attention being a strong contender) that make information widely available in the brain. The present suggestion, however, is much simpler. If (for whatever reason) attention plays some role in determining the contents of visual awareness, perhaps gross motor signals could affect conscious visual experience *by affecting or helping to engage processes of attention*. Thus, recall the case of the post-commisuratomy patient. Perhaps what matters here is not the gross motor signal *itself* (the intention to reach with the left hand) so much as the motor-dependent disposition of *attention*. Where Hurley depicts the case as one in which "the side of motor activity seems to be determining the side of visual awareness" (Hurley 1998, p. 355), we might wonder whether the gross motor intention is more perspicuously seen as engaging an attentional mechanism which in turn, and crucially, modulates conscious awareness.

It is worth pausing to clarify this suggestion.<sup>21</sup> The idea here is *not* that the fine-grained motor signals (the province, let us assume, of the dorsal stream) influence ventral stream activity and thus conscious awareness. Rather, it is that the gross motor *intention* (in this case, simply to move the left hand) results in a disposition of attention that directly influences phenomenal awareness. Such influence seems both intuitively familiar and neuroscientifically plausible. There is plenty of evidence, for example, that the deliberate assignment of attention (even covert attention, with no associated eye movement) modulates the receptive field properties of cells in V1, V2, V4 and MT (see e.g. Motter (1993), Assad and Maunsell (1995)). And PET studies show that shifts of covert attention intentionally directed to specific aspects of a stimulus such as its color, shape and spatial location, results in increased activity in the neuronal groups specialized for processing that type of information (Corbetta, et al (1991) (1993), Haxby, et al (1993)). This kind of evidence is thus distinct from (but compatible with) the idea, embraced by Milner and Goodale, of other processes of selective attention operating entirely *within* the dorsal stream and having no influence on the contents of conscious awareness – see their endorsement

of Rizzolatti, et al's (1994) "premotor" theory of selective spatial attention in Milner and Goodale (1995) p. 183-188.<sup>22</sup> The present suggestion, to recap, is that gross motor intentions engage attentional mechanisms which in turn modify the response characteristics of neuronal groups implicated in conscious awareness.

It is not clear (to me) whether Hurley should regard this as a significant revision. In one sense it would seem to be, since it suggests that the motor signal does not itself help *constitute* the state of visual awareness - rather, it is now depicted as engaging neural mechanisms of attention which *do* seem to play a constitutive role. Against this, however, we must set the fact that what Hurley really aims to show is just the non-instrumental dependence of perception on motor output. If this is defined simply as motor output making a difference, independent of any change in gross sensory input, to the contents of perceptual awareness, the case stands regardless. It does not matter whether the non-instrumental dependence is direct or indirect (going *via* its effects on attentional mechanisms).

It is important in addressing such issues (and thanks to Susan Hurley, personal communication for clarifying this) to distinguish between arguments that attempt to show that motor output makes a constitutive contribution to the fixation of conscious contents and ones that attempt to show that the physical vehicles of the contents (roughly, the most restricted set of inner states sufficient to cause the experience) are *themselves* whole, extended dynamical loops including motor output circuitry. Hurley (1998) attempts both kinds of argument, but merely establishing a non-instrumental role for motor outputs in content fixation cannot itself warrant any conclusions about extended physical vehicles. The present suggestion (concerning the attention-engaging role of gross motor attentions) is best seen as affecting only the issues concerning conscious content fixation. In Hurley's own terms, the idea would be that instead of the gross motor intentions making a constitutive contribution to conscious contents, they affect what she terms the "borderland" of attentional mechanisms. It is the borderland activity itself which then makes the decisive contribution to content.

The main moral, however, is independent of these additional speculations about a bridging role for attention. It is that there is plenty of evidence that visual awareness is action-oriented and non-instrumentally motor-sensitive, but that this is fully compatible with Milner and Goodale's depiction of the dorsal stream as a semi-insulated system for visuo-motor action and of the ventral stream as a semi-insulated system for visual awareness. For the sense in which visual awareness is tied up with action and motor commands is quite distinct from the sense in which the non-conscious dorsal stream "takes care" of action. What influences visual awareness, I have suggested, is a kind of schematic intentional version of motor control: one involving broad motor plans, projects and intentions. While what falls to the dorsal stream is the fine-tuned implementation of these plans. Both systems are thus profoundly motor-oriented. But in different and complementary ways. This complementarity is nicely captured by Goodale himself, who recently suggests that:

[the] interplay between a 'smart' but metrically-challenged ventral stream and a 'dumb' but metrically-accurate dorsal stream is reminiscent of ... what engineers call teleassistance – where a human operator looks at a scene, say the surface of a hostile planet, makes a decision that a particular rock needs to be examined, and then sends a command to pick up the rock to a semi-autonomous robot on the planet's surface.

Goodale (1998) p. 491

Notice, finally, that none of this need imply a return to the old notion of a central executive. What matters is *not* that we identify the full intelligence of the agent with the ventral stream (which would, I think, be a serious mistake: the tele-assistance metaphor misleads in this respect). Rather, what matters is that we recognize the computationally efficient division of labor achieved by using a semi-insulated system for fine-tuned visuomotor action, and a semi-insulated system for visual awareness.

Recognition of such neural division of labor need not (and should not) blind us to the equal importance of large-scale distributed dynamics. It is perfectly possible, for example, to hold that visual awareness itself depends on much more than mere ventral stream activity. It could

involve, for example, complex recurrent dynamics linking multiple cortical and sub-cortical sites. Complex dynamic loops, as suggested by theorists such as Skarda and Freeman (1987), Freeman (1991) (1995), Edelman (1987) and Damasio (1994), may even act as the neural vehicles of many kinds of mental content, and these loops may effectively combine different features into meaningful packages geared to an animal's current needs and goals. Visual awareness may thus reflect current context, projects and memories as well as relevant aspects of ongoing sensory input. What it need not (and probably should not) reflect is the finest-grained detail of visuomotor action systems.

#### **4. Conclusions: Balancing Intimacy and Estrangement.**

The cognitive scientific understanding of perception, awareness and action is, I conclude, likely to turn on the appreciation of an especially complex kind of dynamic balance. It is a balance, as I have argued elsewhere (Clark (1997) ch. 7), between large-scale, multi-component dynamics and pockets of local order and specialization: a balance between *intimacy* and close cooperation on the one hand, and *estrangement* and semi-autonomous specialization on the other. In considering the relations between perception, awareness, and action, we must do simultaneous justice to processes and dynamics of both kinds, and (hardest of all) to the ways in which the two harmonize and interact.

In the specific case we considered – the role of embodied action in visual awareness – there is convincing evidence that perception and action are both intricately intertwined and multiply dissociated! There is clear evidence of fine-tuned action-oriented coding in the dorsal stream. But there is also suggestive evidence that this whole visual stream operates semi-autonomously from the ventrally-dominated systems underlying major aspects of visual awareness. Within the ventral stream itself, however, we find *another* kind of interpenetration of perception and action: the kind stressed by Hurley and involving the influence of gross motor intentions and schematic action plans on conscious visual content. Finally, we must also consider how best to conceptualize the way these two semi-autonomous systems work in harmony so as to yield useful visual awareness of the very world in which we move and act.

The case of visual perception and action thus presents a much more complex problem than the case of rhythmic motion with which we began. In the rhythmic motion case, there was a clean “same-target” conflict between a centralist, localist model of a specific phenomenon and a large-scale, interactionist, dynamical alternative. In the case of visual perception and action, there are multiple phenomena presenting very different explanatory targets, and at least two quite distinct ways in which visual and motor elements may be said to combine and interact.

Let me end, then, by relating all this to some of the bigger issues implied by the overarching theme of “reclaiming cognition”. Recent years have indeed, or so it seems to me, been marked by the emergence of a new kind of science of the mind – one that places embodiment and action at the forefront, and that recognizes the crucial role of distributed dynamics (rather than static symbolic structures) in underpinning human thought and reason. But as with all dramatic shifts in emphasis, there is a concurrent danger – the danger of letting the pendulum swing too far in the opposite direction. For a mature science of the mind needs, somehow, to do *simultaneous* justice to the emergent unity (courtesy of looping webs of dynamical influence) and the frequent sub-systemic estrangement (courtesy of computationally efficient pockets of specialization and insulated functioning) characteristic of biological brains and natural intelligence. Current cognitive scientific research still ends to be drawn to one or other of these poles, oscillating between stress on complex dynamical intimacy and recognition of significant specialization and dissociation. Nature, as ever, contrives to have it both ways and all at once. Finding the models, metaphors and analytic tools necessary to describe and comprehend this unique balancing act is vital if cognition is to be truly *reclaimed*, rather than simply buffeted by another academic tug of war.

---

<sup>1</sup> For a nice - and also ultimately reconciliatory - account, see Bechtel and Richardson (1993), as well as Fontana and Ballati (1999).

<sup>2</sup> The focus of embodiment, in the recent tradition, probably began with Dreyfus (1979), Thompson, Varela and Rosch (1991), and Johnson (1987). Historical precedents clearly include Merleau-Ponty (1942), Heidegger (1927) and, to some extent, Gibson (1979). The specific phrase “embodied, embedded” is due to Haugeland (1998).

<sup>3</sup> Haugeland (1998), Merleau-Ponty (1942), Hatsopoulos (1996), Chiel and Beer (1997), Beer (1995), Hurley (1998) - to name but a few.

---

<sup>4</sup> For a sampling, see Kugler and Turvey (1987), Kelso (1995), Thelen and Smith (1994), Port, Cummins and McCauley (1995), Hatsopoulos and Warren (1996). Historical antecedents include Bernstein (1967).

<sup>5</sup> E.g. Wallace (1981), Gibbon, Church and Meck (1984).

<sup>6</sup> Hatsopoulos (1996), Hatsopoulos and Warren (1996)

<sup>7</sup> Hatsopoulos and Warren (op cit) p. 12

<sup>8</sup> of some cells in the arm/hand region of monkey postcentral gyrus.

<sup>9</sup> For such arguments, see Hutchins (1995), Clark (1995), Clark and Chalmers (1998), Dennett (1995) ch. 11,12.

<sup>10</sup> In this vein, Koch and Braun recently wrote:

We believe that further progress in understanding visual, and indeed all forms of, awareness will only come from precisely locating the NCC [neural correlates of consciousness] to specific brain areas and local populations.

Koch and Braun (1998) p. 170

Crick and Koch (1995) thus argue, for example, that the NCC for vision includes only those visual neurons that directly project into frontal cortex. The account thus encompasses both the idea that only a subset of cortical neurons 'matter' *and* the famous idea (Crick and Koch (1990) - reprinted as Crick and Koch (1997)) that binding by synchronous activation is a crucial part of the mechanism by which a specific content then becomes conscious.

<sup>11</sup> Farah (1997) draws exactly this distinction, contrasting accounts which depict consciousness as "the privileged role of particular brain systems" with ones which stress "states of integration between brain systems". Op cit p. 203-215. As examples of such integrative accounts Farah cites Kinsbourne (1998), Crick and Koch (1990), and Damasio (1990). We might add Edelman (1992) and even Dennett's (1993) account of consciousness as "cerebral celebrity".

<sup>12</sup> See Van Essen and De Yoe (1995), Knierim and Van Essen (1992), Wurtz and Mohler (1976).

<sup>13</sup> See also Clark (1997) on "Continuous reciprocal causation" and Jarvilehto (1998).

<sup>14</sup> The scare quotes are included because these distinctions, in such accounts, are frequently themselves called into question.

<sup>15</sup> Susan Hurley (1998) p. 183, speaks of the "twisted rope" metaphor of neural processing.

<sup>16</sup> Hurley herself addresses this question in a long footnote on pages 183-184 of Hurley (1998). In the footnote Hurley accepts much of the Milner & Goodale story, but questions the "overlay" which assimilates the dorsal/ventral systems to systems dedicated to action/conscious perception, pointing out the ventral functions "include responses and hence actions-it's just that these are much more flexible and cognitively immediated" (op cit). The account I develop is compatible with this, but goes further in attempting to distinguish two classes of motor output signal, only one of which is plausibly implicated in the constitution of conscious visual awareness.

<sup>17</sup> Milner and Goodale comment, revealingly, that "understanding these interactions would take us some way towards answering what is one of the central questions in modern neuroscience: how is sensory information transformed into purposeful acts" (Milner and Goodale 1995, p. 202).

<sup>18</sup> See e.g. Mack and Rock 1998, pp. 232-235.

<sup>19</sup> Note that the claim is that attention is *necessary* for a stimulus to become consciously perceived. It need not - and probably should not - be counted as sufficient.

<sup>20</sup> Thanks to Jesse Prinz for helping to convince me of this.

<sup>21</sup> Thanks to an anonymous referee for pressing me on this point.

<sup>22</sup> This kind of attention-invoking story is also clearly compatible with Milner and Goodale's further suggestion that conscious awareness of visual content requires both ventral stream coding *and* a sharpening effect due to "spatial gating processes known to be active during selective attention" (Milner and Goodale 1998, p.7.)