The scope and limits of enactive approaches to visual experience

Pierre Jacob
Institut Jean Nicod, UMR 8126, EHESS/ENS-DEC/CNRS, F 75007 Paris, France;
e-mail: jacob@ehess.fr

Abstract. I pursue here three related aims. First, I criticise some of the metaphysical claims made on behalf of the so-called 'enactive' approach to visual experience. Secondly, I explain why the enactive view of visual experience is hard to square with the evidence in favour of the two-visual-systems model of human vision. Finally, I explore one possible way to develop the 'pre-emptive perception' framework and explain why, contrary to first appearances, some of the fundamental discoveries of brain mechanisms, whose function might be to underlie pre-emptive perception, do not really support the enactive approach to visual experience.

1 Introduction

According to the classical (or orthodox) 'computational representational' paradigm in cognitive science, well-exemplified by David Marr's (1982) approach, visual perception results from the brain's computing and transforming a series of mental representations in response to visual inputs hitting the retina. In recent years, this paradigm has been criticised by supporters of such alternatives as eg 'embodied cognition' and the so-called 'enactive' approach to visual experience. What advocates of both enactivism and embodied cognition object to is the view of the mind as a disembodied Turing machine manipulating formal symbols according to syntactic rules. In particular, the assumption that visual perception involves the creation and transformation of internal mental representations has been questioned on the grounds that the perceived world itself can serve "as its best own model" (Rodney Brook) or as an "outside memory" (Kevin O'Regan). As the philosopher Alva Noë (2004) has recently put it on behalf of the enactive approach to visual perception:

> "There's no need to re-present the world on one's own internal memory drive. Off-loading internal processing onto the world simplifies our cognitive lives and makes good engineering and evolutionary sense." (page 50)

Now, as the title of the 2005 Hanse-Wissenschaftskolleg Conference testifies, a new intriguing label has recently been introduced into the science of vision and the philosophy of perception by the vision scientist Ivan Bodis-Wollner (2005): "pre-emptive perception". How could perception be 'pre-emptive' and what could it pre-empt? I myself am pretty old-fashioned and I don't think that any really convincing reasons have been provided for giving up the classical computational representational approach to the puzzles of visual perception.[1] So, in the present paper, I have three related aims: first, I want to discuss some of the metaphysical claims made on behalf of the enactive approach to visual experience. Secondly, I want to explain why I think that the enactive view of visual experience is hard to square with the evidence in favour of the two-visual-systems model of human vision. Finally, I want to explore one possible way to develop the 'pre-emptive perception' framework and explain why, contrary

---

[1] As I will argue here, I don't think that psychophysical experiments on change blindness provide reasons for giving up the computational-and-representational framework. Nor do I think that results from cognitive neuroscience force us to do so either.

to first appearances, some of the fundamental discoveries of brain mechanisms, whose function might be to underlie 'pre-emptive perception', do not really support the enactive approach to visual experience either.

2 What is enaction?

So-called 'enactive' approaches to visual perception and visual experience are at present hotly debated in cognitive science and the philosophy of perception (see eg O'Regan and Noë 2001). The word 'enaction' was introduced into cognitive science by the late Francisco Varela. According to the English Oxford Dictionary (EOD), the verb 'to enact' means both "to ordain (a decree)" and "to interpret a play".

Whatever exactly the word 'enaction' is supposed to mean in the context of scientific and philosophical discussions of visual perception and visual experience, there is little doubt, I think, that enactive approaches to visual perception are instances of *motor* theories and that they emphasise the contribution of planning and execution of action and bodily movements to visual perception and visual experience. As O'Regan and Noë (2001) and Noë (2004) emphasise on behalf of the enactive conception, perception is not something that happens to one, it is something one does. Perception is intrinsically active: it is the skillful bodily exploratory activity of one's environment. What one perceives—the content of one's perceptual experience—is (partly) constituted (or determined) by one's (implicit) knowledge of the sensory consequences of one's movements.

Now in the relevant sense, 'constitution' is a word that belongs to the philosophical jargon. To say that the content of one's visual experience of eg a tomato is 'constituted' by one's sensorimotor knowledge of the sensory consequences of one's own actions relative to the tomato is to say that one's sensorimotor knowledge of the sensory consequences of one's actions directed towards the tomato is both necessary and sufficient for one to undergo a visual experience of a tomato. For example, consider the visual experience of a tomato resting on a plate. Although only part of the surface of the tomato reflects photons onto one's retina, nonetheless, as Noë (2004) puts it, one's perceptual sense of the 'presence' of the hidden and invisible surface of a tomato arises from the fact that one (implicitly) knows that if one were to manipulate the tomato with one's hand and turn it around, one would enjoy a visual experience of the parts of the tomato's surface that one is not currently seeing (ie that are not at that instant reflecting impinging photons onto one's retina).[2]

If the enactive account claimed that perceiving an object is just to act on it, then it would, I think, run the risk both of being inconsistent with all the evidence in favour of the two-visual-systems model of human vision (about which see section 3) and of being committed to an implausible form of behaviourism as well. But the enactive account does not claim that perceiving an object consists of acting on it. The enactive account claims that to perceive an object consists of having some (implicit) sensorimotor *knowledge* of the sensory consequences of one's own actions directed towards the object. Although the enactive account is more subtle than straightforward behaviourism was, it is nonetheless threatened by some kind of instability—if not circularity. Unless one were able to sort it out from neighbouring distractors, how could one plan to act on some target?

---

[2] Some of the problems with the enactive account arise from the lack of clarity of what counts as possessing or lacking "the (implicit) sensorimotor knowledge of the sensory consequences of one's own actions". For example, does a person incapable of either walking and/or moving upper limbs, as a result of a stroke, lack the knowledge in question? If this person was deprived of the ability to grasp and manipulate a tomato and knew that someone else could do so in front of him/her, does the theory predict that the paralysed person would enjoy (or be deprived of) a normal percept of the tomato? For further queries, see my discussion of the double dissociation between visual form agnosia and optic ataxia in section 3.

For someone to know (implicitly or explicitly) how to act on an object—for anyone to know the sensory consequences of one's own actions upon some object—one must be able to independently track the object or distinguish the putative target of one's action from other surrounding objects and from the background. On the enactive account, one's visual experience of an object is constituted by one's knowledge of which sensory consequences one's actions would have upon the perceived object. But, now, if one's knowledge of the sensory consequences of one's actions upon the object requires some independent and prior tracking or sensory detection of the object, then the enactive account seems plainly threatened by circularity. At the very least, enactivism does not seem to afford a constitutive account of the content of visual experience.

Elsewhere, I have already expressed some of my reservations about several features of the enactive conception of visual experience (Jacob 2006). In particular, I have taken issue with the claim made by several advocates of the enactive account of visual experience that experimental work on change blindness is inconsistent with the classical computational representational approach to visual perception. The term 'change blindness' refers to the surprising experimental discovery that healthy human adults fail to notice important changes in a rich visual scene when they are presented with repeated alternations between an original scene and a modified version separated by a brief blank display and when normal cues of change between two scenes are skillfully concealed from observers by experimenters.[3]

As Dretske (2004) and Block (2005a) have put it, what change blindness does show is that healthy human adults may fail to *notice differences* between two scenes or two displays that are produced by a change. Now, one could not notice a difference between two scenes (caused by a change) unless one made a *comparison* between the two scenes. So what change blindness shows is that healthy human adults can fail to make a comparison between two distinct scenes. Granted, one could fail to make such a comparison if one lacked a (detailed) representation of each scene. However, as noticed by Simons and Rensink (2005), there are many alternative explanations of subjects' failures to make the appropriate comparisons that are consistent with the fact that they do form (detailed) representations of the two displays. For example, the detailed representation of the stimulus perceived at t_1 could be encoded in a neural pathway unavailable for comparison with a representation of a stimulus perceived at t_2. Alternatively, the comparison process could fail because it is too slow relative to the temporal interval allowed by the experimenters. Finally, some recent experimental work reported by Lamme (2003) and discussed by Block (2005b) strongly suggests that change blindness may arise from limitations, not of visual awareness per se, but rather of mechanisms of visual *attention* (and short-term memory). Therefore, I do not think that the work on change blindness demonstrates that vision cannot consist in forming detailed internal representations of the world.

Furthermore, Noë's (2004) appeal to work on change blindness plays a crucial role in his argument in favour of a puzzling metaphysical view, which, following Clark and Chalmers (1998), he calls "vehicle externalism". On behalf of what is often called "embodied cognition", Clark and Chalmers (1998) and Rowlands (1999) have argued that props (eg computers and books) are constitutive (ie part of the individuative conditions) of the contents of some thoughts, on the ground that they are inextricably bound with some thought processes. However, using paper and pencil in order to prove some arithmetical theorem does not make paper and pencil part and parcel of the 'semantic' content of the theorem. Nor do two proofs of one and the same arithmetical

---

[3] Similar experimental results have been obtained for changes made during saccadic eye movement, eye blinks, so-called 'mud splashes', and brief occlusions [see O'Regan (2000), and Simons and Rensink (2005) for a review].

theorem, one with paper and pencil, the other with a computer, confer different contents to two versions of the same theorem.

Now, according to Noë's (2004, ch. 7) version of 'vehicle externalism', one's visual experiences do not supervene on internal states of one's brain or, as philosophers sometimes like to put it, one's visual experiences are not *in* one's head. To claim that one's visual experiences do not supervene on one's brain is to deny that for any change in one's visual experience to occur, some change or other must occur in one's brain (or visual system).

Noë (2004) offers two conspiring reasons in favour of applying vehicle externalism to the content of one's visual experiences. First of all, some visual processes involve the use of instruments (eg telescopes and microscopes) that are not themselves parts of internal states of one's brain. But, from the fact that some perceptual processes may involve the use of an instrument, it would only follow that the content of one's visual experience fails to supervene on internal states of one's brain if the contribution of an instrument to the perceptual process failed to have a causal effect upon the internal states of one's brain. But why believe that, when one looks through eg a microscope, the state of one's visual system is not causally affected by the interaction with the microscope?[4]

Secondly, Noë (2004) argues that experiments on change blindness establish that the content of one's visual experiences is *virtual*. According to his interpretation, experiments on change blindness are supposed to conclusively show that our experience of the details of a visual scene does not arise from a detailed visual internal representation. The detail belongs, so to speak, to the outside world. It is not intrinsic to our visual representation of it. We implicitly know that the detail is available (ie virtual) and where to retrieve it if needed.

However, as I have already argued, I do not think that the experiments on change blindness show that the content of visual experience is virtual in the required sense. Furthermore, as Richard Gregory has observed (personal communication),[5] from the fact that a visual representation of a perceived scene may be less detailed than the represented scene, it hardly follows that one's visual experience does not arise from a mental representation of the scene, let alone that one's visual experiences do not supervene upon the internal states of one's brain (or that they lie *outside* of one's head). The fact that some features of a visible display (eg Mt Sainte-Victoire in the South of France) are not represented in eg a Cézanne painting shows that the representation (the painting) differs from what it represents (ie Mt Sainte-Victoire itself). It is quite conceivable that there are details in the Mt Sainte-Victoire itself that are not represented in the Cézanne painting. But it would be odd to conclude, on these grounds, that the depicted content of a Cézanne painting is virtual and/or that it is not a property of the canvas (a physical object). Similarly, the fact that an individual's mental representation differs in some respects from, or lacks some of the features of, what it represents does not show that the mental representation is not a state of the individual's brain.

3 Enaction and the two-visual-systems model
As I said above, according to the enactive conception, the content of one's visual experience is (partly) constituted by one's (implicit) sensorimotor knowledge of the sensory consequences of one's actions. For example, the content of one's visual experience of

---

[4] Block (2005a) makes similar points.
[5] As he puts it in an e-mail message of 12.01.2006: "The outside world has more details than are represented (as in a painting or a fuzzy photograph)—but this does not reflect the status of representation. It doesn't make it in the brain or out of it. No representation can be identical to reality—or it would be reality."

a tomato is (partly) constituted by one's sensorimotor knowledge that, if one were to grasp the tomato and turn it upside down, then one could be able to see parts of the tomato's surface that are currently not visible. I argued earlier that the enactive account is threatened by a circularity problem, to which the two-visual-systems model of human vision is clearly immune. On the two-visual-systems model, the selection of the target (eg a tomato) in an act of prehension is a perceptual process based on such perceptual cues as the target's colour, relative shape, relative size, and texture. The visuomotor transformation whereby the agent reaches and grasps the target is an automatic process. Whereas the selection of the target from a set of distractors requires that the target's location be represented in an allocentric frame of reference, the success of the visuomotor transformation requires that it be represented in an egocentric frame of reference. How the visual system converts the representation of the location of a target from one coordinate system into another is a straightforward empirical and computational problem.[6] But on this view, the sensorimotor knowledge that enables an agent to efficiently reach, grasp, and manipulate a tomato is not constitutive of the content of the agent's visual experience of the target (of his/her action).

The question I now want to consider is whether the enactive conception of perception is compatible with the evidence in favour of the two-visual-systems model of human vision, according to which humans can visually process objects for two distinct purposes: for perception and for the visuomotor control of action.[7] I find the evidence in favour of the two-visual-systems model of human vision quite compelling. The evidence comes from electrophysiological recordings in the brain of non-human primates, from double dissociations exemplified by brain-lesioned human patients, and from psychophysical experiments on healthy human adults. Here, I will merely draw attention to some key results from three scientific areas: the neuropsychological examination of brain-lesioned human patients, psychophysical experiments involving illusory stimuli, and some recent experiments on so-called scale errors in 2-year-old humans.

The basic relevant information from neuropsychology is the double dissociation between visual-form (apperceptive) agnosic patients and optic ataxic patients. Whereas visual-form agnosia results from a lesion in the infero-temporal pathway (the so-called ventral stream of the visual cortex), optic ataxia results from a lesion in the superior parietal lobe (in the so-called dorsal stream). Visual form apperceptive agnosic patients are deeply impaired in the recognition of the shapes of visually presented objects. However, patient DF [who has been studied in depth by Milner et al (1991), Goodale et al (1991), Goodale (1995), James et al (2003), and Milner and Goodale (1995)] exhibits surprising residual visuomotor capacities. For example, although she cannot visually compute the shape, size, and orientation of an object for the purpose of recognising and/or identifying it, she can grasp it with almost normal precision grip. Although she cannot scale the distance between her thumb and index finger if the perceptual task is to indicate her estimate of the size of an object, she can do so if the task is to reach and grasp it. Conversely, optic ataxic patient AT [studied by Jeannerod et al (1994) and Milner et al (1999)] is deeply impaired in visuomotor tasks of reaching and grasping objects. However, her ability to recognise and/or identify the shapes of visually presented objects is unimpaired. In particular, unlike patient DF, whereas she fails to reach and grasp an object, patient AT can scale the distance between her thumb and index finger if the perceptual task is to indicate her estimate of the object's size.

[6] About which see next section.
[7] For a detailed discussion of the evidence from electrophysiology, neuropsychology, and psychophysics, see Milner and Goodale (1995) and, for a slightly different account, Jacob and Jeannerod (2003), Jacob (2005), and Jeannerod and Jacob (2005). See also Jacob (2006) and Block (2005a), with whose conclusions I concur. For corroborating arguments, see Mahon and Caramazza (2005).

On the one hand, it seems as if the impairment of visual-form agnosic patients shows that the implicit sensorimotor knowledge of the sensory consequences of one's own actions towards an object is not *sufficient* for visual perception. On the other hand, it seems as if the impairment of optic ataxic patients shows that the implicit sensorimotor knowledge of the sensory consequences of one's own actions towards an object is not *necessary* for visual perception. So the question is: how should the enactive theorist respond to this neuropsychological dissociation between the perceptual processing and the visuomotor processing of an object's visual attributes according to the task? Perhaps the enactive theorist might say that knowing how to reach and grasp an object is not relevant to the visual experience of an object's size, shape, and orientation. Perhaps it is not. But if so, then the question arises: implicit knowledge of which other sensorimotor contingency (or contingencies) is relevant to (necessary and/or sufficient for) the visual experience of an object's size, shape, and orientation?

To save space, I shall discuss only one psychophysical experiment exemplifying the dissociation in healthy human adults between perceptual and motor responses elicited by one and the same illusory stimulus—the Titchener (or Ebbinghaus) circles illusion, where a three-dimensional (3-D) disk surrounded by an annulus of circles smaller than it seems larger than a 3-D disk of the very same diameter surrounded by an annulus of circles larger than it.[8] Haffenden et al (2001) noticed that in traditional presentations of the Titchener circles illusion, the gap between the central 3-D disk and the surrounding annulus is 3 mm when the surrounding circles are smaller than the disk and 11 mm when the surrounding circles are larger than the disk. Hence, they presented subjects with three Titchener disk displays one at a time, two of which were the traditional displays, and the third involved an annulus of smaller circles located at a gap of 11 mm. In other words, the third condition (called "adjusted small" by Haffenden et al 2001) was a hybrid of the two traditional conditions (see figure 1). In each of the three conditions, subjects were requested either to grasp the central disk between thumb and index finger or to show their estimate of the diameter of the central disk by scaling the distance between thumb and index finger. As Haffenden et al (2001) measured the distance between thumb and index finger in both tasks, they found that, whereas the perceptual response to the display of the third condition was very much like the perceptual response to the presentation of the disk surrounded by the smaller circles with a 3 mm gap, the motor response to the display of the third condition was very much like the motor response to the presentation of the disk surrounded by the larger circles with an 11 mm gap. In other words, they found a dissociation between the perceptual and the motor responses to the display of the third condition.[9]

In my view, this experiment raises the following puzzle for the enactive account of visual perception. When a subject estimates the size of the disk by scaling the distance between the thumb and the index finger, she/he presumably uses the very same implicit knowledge of the sensory consequences of the finger movement as when she/he reaches and grasps the disk. After all, only in reaching and grasping an object is an agent acting onto an object. If this is true, then how can the knowledge of the

[8] The idea of replacing a 2-D central circle by a 3-D disk in order to investigate the differences between the perceptual and the motor responses to displays of the Titchener illusion was first exploited by Aglioti et al (1995). For methodological discussions, see Franz et al (2000, 2001) and Jacob and Jeannerod (2003, ch. 4).
[9] Most, if not all, of the dissociations discussed by advocates of the two-visual-systems model of human vision are such that the perceptual system is subject to some illusion to which the visuomotor system is relatively immune. By contrast, Hauser (2003) discusses developmental dissociations exhibiting the converse pattern, in which the perceptual system provides the right answer and the motor system does not.
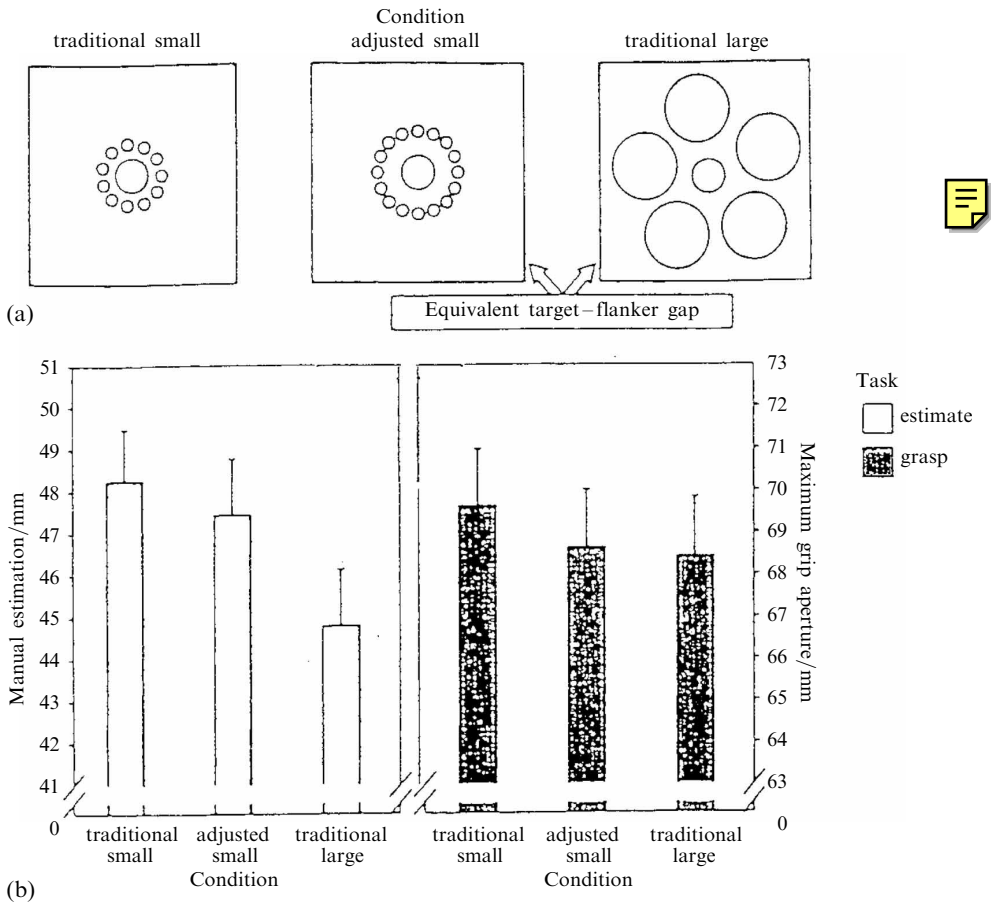
Figure 1. Titchener circles illusion. (a) In the leftmost square there is a small distance between central disk and annulus of smaller circles in the standard Titchener circles illusion. In the right-most square there is a large distance between central disk and annulus of larger circles in the standard Titchener circles illusion. In the central square there is a large distance between central disk and annulus of smaller circles in the nonstandard Titchener circles illusion. (b) White bars show manual estimation for each of the three conditions displayed separately: traditional small, traditional large, and adjusted small. Black bars show maximal grip aperture in grasping task in each of the three conditions. The experiment shows dissociation between manual estimation and grasping task: manual estimation of adjusted small is very much like manual estimation of traditional small; but grasping of adjusted small is very much like grasping of traditional large. Reproduced by permission from Haffenden A M, Schiff K C, and Goodale M A, 2001 *Current Biology* **11** 177 – 181.

sensorimotor consequences of finger movements explain the dissociation between the perceptual and the motor tasks in the third condition? If the advocate of the enactive approach were to rule out the sensorimotor knowledge of grasping an object as irrele-vant to perceiving the object's size, then, I take it, he or she would simply leave the above dissociation unexplained.

Finally, let me briefly discuss recent work by DeLoache et al (2004) on scale errors in 2-year-olds, which shows an interesting dissociation between toddlers' planning and execution of actions. Children aged 18 to 30 months were first brought into a room containing three large toys (such as a toy car, a slide, and a chair) with which they played. Next, they were escorted out of the room. Finally, they were brought back into the room, which now contained miniature replicas of the three kinds of large toys. On one hand, DeLoache et al (2004) report that children attempted to act on

the miniature toys in a manner appropriate to their larger models, but inappropriate to their miniature replicas. On the other hand, they found that, in spite of the fact that children selected actions that were appropriate to larger models and inappropriate to miniature replicas, children's execution of these actions showed accurate adaptation of their movements to the actual size of the miniature replicas. These so-called 'scale errors' seem to show that the visual perception of a (miniature) object evokes in children a representation of the larger model and activates a mental representation of an action plan appropriate to the larger model, with which the child has been used to interact.

DeLoache et al (2004) surmise that, in virtue of the immaturity of their frontal lobes, the children are unable to inhibit the execution of the inappropriately planned action. However, unlike the planning of the children's actions, the finely tuned online execution of their bodily movements is not controlled by the perceptual and/or memorised representation of the larger models. It is based on a veridical appreciation of the size of the miniature replicas. This finding raises several interesting theoretical problems: it provides evidence that, in toddlers' brains, information about the identity of a perceived object (processed in one stream of the visual cortex) may fail to be integrated with information about the object's size (processed in the other stream). Glover (2004) argues that it reflects his own distinction between the planning and the online execution of actions. As acknowledged by DeLoache et al (2004), it could also reflect the duality between the visual selection of a target for action and the detailed visual control of actions upon the chosen target, emphasised by Milner and Goodale (1995). Whatever the exact explanation of DeLoache et al's (2004) finding, it is hard to see how the enactive theorist could accommodate this dissociation, because it suggests that the perceptual content of an agent's visual experience of the size of an object cannot be constituted by his/her implicit knowledge of the sensory consequences of using size information for the purpose of acting upon the object.

When considering whether the enactive account is consistent with the evidence in favour of the two-visual-systems model of human vision, Noë (2004, pp 12, 19) quickly dismisses the question by distinguishing the claim, which he takes to be *constitutive* of the enactive conception, that visual perception is made possible by the perceiver's sensorimotor "practical grasp of the way sensory stimulation varies as the perceiver moves" from a putatively distinct presumably empirical claim, which he rejects, according to which "perception is *for* acting or for guiding action". No doubt, the two-visual-systems hypothesis is an empirical hypothesis that should not be treated as an infallible dogma.

However, I am not convinced that Noë can reconcile the enactive conception with the evidence in favour of the two-visual-systems hypothesis by conferring to his own appeal to one's sensorimotor knowledge of the sensory consequences of one's own actions the status of what philosophers call a 'constitutive' claim, ie a claim supposed to express a conceptual truth. Nor is it always clear that Noë (2004) really rejects the view that perceptual experience guides action, as when he writes (eg on page 90) that "there is no question that experience can and does guide movement, but it is not the business of the enactive view to emphasize this humdrum fact". The evidence for the two-visual-systems hypothesis is precisely what would lead one to doubt at least the strong assertion that the function of perceptual experience is to guide movement.

On several occasions, Noë endorses consequences of the view that the function of visual perception is to serve action. For example, in the footsteps of Poincaré's famous view that "to localize an object simply means to represent to oneself the movements that would be necessary to reach it" (quoted by Noë 2004, p. 75), Noë seems to take the view that localising an object in some egocentric frame of reference (centred on the axis of the agent's body) is a necessary condition for one to experience what

Noë calls its "perspectival" properties or P-properties (for short). For example, a circular coin may both have the shape property of being circular and look elliptical. Its looking elliptical is one of the coin's properties which Noë calls its P-properties. For example, he writes:

> "The plate looks elliptical to me because, to indicate its shape, I can (and indeed, in some sense, *must*) move my hand in a characteristic manner. That is, to experience a thing as elliptical is precisely to experience it as occupying a particular region in one's egocentric, sensorimotor space ... The P-size of an object is given precisely as that region to which I, for example, would point or reach (etc.) if I wished to direct myself to the object". (p. 89)

On one hand, Noë's account of the P-shape of the plate smacks of behaviourism: my hand movements may reveal or manifest my visual experience of the P-shape of the plate (eg to others), but the former *does not* explain the latter. On the other hand, the human visual system can represent the spatial location of an object relative to several different frames of reference, eg relative to an egocentric frame of reference or relative to an allocentric frame of reference (centred on some constituent of the visual array).[10] An object's location must presumably be represented in some egocentric frame of reference (centred on part of the agent's body) for the agent to act upon it (eg to reach and grasp it). But it is by no means obvious that an object's location must be represented in an egocentric frame of reference for an observer to visually experience either the object's P-shape or its P-size. Indeed, unless visual agnosic patient DF were able to represent an object's location in an egocentric frame of reference, she would not be able to grasp it efficiently. However, the fact that she cannot visually recognise the shape of objects suggests that representing the location of an object in an egocentric frame of reference is not sufficient to enable an agent to experience an object's shape, let alone its P-shape. If it is part of the recognition of an object's shape to experience what Noë calls its P-shape, then arguably representing the location of an object in egocentric coordinates is not sufficient for the experience of the object's P-shape.

4 Why pre-emption is not enaction

As I said in section 1, at first blush the idea of 'pre-emptive perception' sounds odd. But, on reflection, it may convey an important insight. According to the authority of the EOD, the word 'pre-emption' would seem to apply to an action in the course of which some agent A intentionally strikes another agent B in order to deprive B of the ability to strike A. In other words, it seems as if no action by agent A could count as 'pre-emptive' unless A had the cognitive capacity to recognise B's putative intention to strike him and to entertain a higher-order reflexive intention whose content would involve, among other things, the representation of B's intention to strike A, ie including a reference to oneself (as a putative target of B's attack).

Clearly, however, the primary goal of the 'pre-emptive perception' framework is not intended to apply to the perception of the *actions* of one's conspecifics. Surely, it is intended to apply primarily to basic cases of the visual perception of inanimate objects. Of course, photons reflected by ordinary inanimate objects do strike an observer's retina. But, clearly,

---

[10] In contemporary philosophy and cognitive science, the word 'egocentricity' is sometimes used to capture the context-dependence of demonstrative and indexical *thoughts* (eg expressed by 'here'), which have genuine *conceptual* content. It is also used to refer to a property of a particular kind of frame of reference relative to which the visual system represents the spatial position of an object. On one hand, egocentric frames of reference, which are viewer-centred, contrast with object-centred or allocentric frames of reference. On the other hand, arguably visual representations of objects based on an egocentric frame of reference lack conceptual content.

we don't want to burden the 'pre-emptive perception' framework with the assumptions that photons and/or the inanimate objects that reflects them onto an observer's retina are intentional agents, let alone that an observer might intend to deprive either photons or the inanimate objects that reflect them of their ability to (intentionally) strike him/her.

Nor, I think we need do so. There is much evidence that there are brain mechanisms in many different species that enable an animal to keep track of its own self-generated movements, for both motor and sensory purposes (Wurtz and Sommer 2004). For the purpose of coordinating complex motor sequences, an animal's motor system needs to be constantly updated about the immediately executed motor commands before it can execute the next steps. For the purpose of visual processing, an animal must be able to discriminate sensory changes that are exogenously caused in its environment from those that arise as a result of its own self-generated movements: unlike the former, the latter are predictable and should, therefore, be less surprising. In the rest of this paper, I will emphasise the contribution of so-called 'corollary discharge' information to sensory processing, at the expense of the contribution to the organisation of motor behaviour. As I will interpret it, at the core of the 'pre-emptive perception' programme is the idea that much visual perception is facilitated by the existence of specialised brain mechanisms that anticipate the result of a self-generated movement based on corollary discharge information about the impending instruction to move. In other words, as I will construe it, the 'pre-emptive perception' programme emphasises the contribution of an animal's ability to keep track of its own self-generated movements to visual processing. If so, then the task of these specialised brain mechanisms is not to pre-empt another intentional agent to strike. Rather, their task is to anticipate the animal's launching of self-generated movements required for visual perception to occur. But the question arises: is 'pre-emptive perception' just another name for 'enaction'? To this question, I will offer a firmly negative answer.

Primates explore their visual environment by means of rapid (ballistic) so-called 'saccadic' eye movements. Ever since the great German nineteenth-century physiologist Helmholtz, vision scientists have been exercised by the puzzle of how the primate brain is able to discriminate the visual signals produced on the retina respectively by the motion of some external moving object and by the movements of the eye. From a biological point of view, three clues are available (cf Wurtz and Sommer 2004). First, whereas the motion of a moving object stands in contrast with a still background, the whole visual field moves as a result of a saccadic eye movement. Second, unlike the motion of a moving object as such, eye movement produces proprioceptive information about eye-muscle contraction. Third, as hypothesised by Helmholtz, there is much evidence that, when the brain generates a saccadic eye movement, it also keeps track of its own motor command by means of a so-called 'corollary discharge' or 'efference copy' of the motor command. Much neuroscientific work has established the existence of corollary-discharge signals in the primate visual oculomotor system. Furthermore, an anatomical pathway conveying corollary-discharge information about saccadic commands has been identified in the primate visual system (Wurtz and Sommer 2004).

As Sommer and Wurtz (2002) and Wurtz and Sommer (2004) usefully emphasise, for neurons to convey corollary-discharge information, they must satisfy the twofold constraint: they must originate from a brain area known to generate movements (in this case eye movements) and they must "project upstream, away from motor neurons, instead of downstream, toward motor neurons" (Sommer and Wurtz 2002, p. 1480). Thus, they recorded single neurons in the monkey mediodorsal thalamus (MD), known to receive projections from the (intermediate layer of the) superior colliculus (SC) and to send projections onto the region of the frontal lobe known as the 'frontal eye field' (FEF) (Sommer and Wurtz 2002).

In a first round of experiments, they showed that neurons in MD have increased activity just before saccade generation. Then, they used an elegant design in which the monkey was engaged in a double-step task involving two successive saccades in total darkness, such that the latter could not be correctly executed unless the monkey had information about the former's execution. In the course of the whole task, the monkey looked at a single fixation point. First, the monkey performed horizontal saccade S1 to stimulus T1 flashed $10°$ away from the fixation point. Second, the monkey performed vertical saccade S2 to stimulus T2 located right on top of T1. The monkey's ability to correctly perform vertical saccade S2 from T1 to T2 depended crucially on corollary-discharge information about the landing site of the first saccade. Sommer and Wurtz then reversibly inactivated neurons in MD by injecting muscimol at the sites where they had previously recorded MD neurons that showed enhanced activity prior to saccade generation. Comparison of the saccadic sequences made before and after inactivation by muscimol injection revealed a significant contraversive shift of the second saccade, as if the monkey had incorrectly planned the execution of the second saccade, not from the landing site of the first saccade, but from the fixation point. Before inactivation, S2 went straight up from T1 to T2, making a right angle with S1. After inactivation, S2 deviated to the right, no longer making the required right angle with S1 for the eye to land on T2.

Interestingly, inactivation of MD neurons by injection of muscimol impaired the second of a sequence of two related saccadic eye movements, not single saccades. Sommer and Wurtz (2002) convincingly argue that the contraversive shift of the second saccade is evidence that the mechanism for corollary discharge has been impaired by muscimol injection. They further argue that MD neurons belong to an upstream pathway from SC neurons towards FEF neurons, carrying corollary-discharge information about saccadic eye movement.

In an earlier paper, Duhamel et al (1992) reported a surprising (if not a paradoxical) finding. They recorded single neurons from the monkey lateral intraparietal area (LIP) both while the animal was fixating and while it was making a saccade from one fixation point to the next. During fixation, a given LIP neuron encodes a stimulus at a certain retinal location (in retino-centred coordinates). If the animal produces a saccade to a new fixation point, the same LIP neuron shifts its so-called 'receptive field' in retinocentric coordinates; it responds to a stimulus at a new retinal location and it stops responding to the stimulus at the previous retinal location (at which it responded prior to the saccade). Now, in the fixation condition, Duhamel et al (1992) found that LIP neurons respond to a visual stimulus 70 ms after its onset. In the saccade condition, the change in fixation target was synchronised with the occurrence of a new visual stimulus, which was positioned at the new expected location of the receptive field of the recorded LIP neuron (after the saccade). To their surprise, in the saccade condition, Duhamel et al (1992) found that, far from responding to the visual stimulus at the location of its new receptive field 70 ms after the eye movement had brought the stimulus within the actual borders of the receptive field of the neuron, the LIP neuron started to discharge some 150 ms earlier, that is 80 ms *before* the beginning of the saccade.

In other words, they discovered that LIP neurons are able to shift their receptive field to their expected position 80 ms before the eye movement actually takes place. Thus, LIP neurons perform a 'predictive remapping' of their receptive field in advance of a self-generated eye movement. As Duhamel et al put it, LIP neurons anticipate the retinal consequences of intended eye movements. Then, the subsequent saccadic eye movement restores the neuron's shifted receptive field to its proper retinal location, enabling the neuron to respond to afferent visual information. Both the impending saccadic eye movement and the occurrence of the visual stimulus at the expected

location of the receptive field of the neuron are required for it to achieve its predictive mapping. The mere presence of the visual stimulus at the appropriate location will not trigger the anticipatory activity of the LIP neuron if the monkey is not planning the saccade. Nor is the intention to execute a saccade sufficient in the absence of the visual stimulus. Duhamel et al (1992) argue that, unless LIP neurons had corollary-discharge information about the saccadic motor command prior to the execution of the saccadic eye movement, they could not perform the predictive remapping of their receptive field.

Arguably, both the discovery, reported by Wurtz and Sommer (2004), of the pathway from SC to FEF for conveying corollary-discharge information about saccadic eye movements, and the discovery of the anticipatory activity of LIP neurons, reported by Duhamel et al (1992), are discoveries of mechanisms that, following Bodis-Wollner (2005), deserve to be called mechanisms of 'pre-emptive perception'. Now, the question arises whether this discovery supports the enactive approach to visual experience, according to which the content of one's perceptual experience arises from (or is constituted by) one's sensorimotor knowledge of the sensory consequences of one's own body movements with respect to perceived objects. I shall now explain why I think that the discovery of mechanisms of 'pre-emptive perception' such as the anticipatory activity of LIP neurons fails to provide support to the enactive account of visual experience.

As my previous discussion of the anticipatory activity of LIP neurons, following Duhamel et al's (1992) findings, made clear, LIP neurons have retinocentric receptive fields. They discharge in response to a stimulus at a retinal location and they achieve a predictive remapping of their receptive field at a new expected retinal location. In other words, a LIP neuron knows the retinal location of a given visual stimulus during eye fixation and it has the amazing ability to predict the retinal location of a new visual stimulus 80 ms before an intended eye movement that will enable it to receive incoming information from this new retinal location. Although the visual representation of the spatial position of any visually perceived object must start with some representation of its spatial position in retinocentred coordinates, we presumably *do not* visually perceive or consciously experience the spatial location of a perceived object in *retinocentred* coordinates. Representing a visual stimulus at a retinal location is a necessary step in the visual perception of a distal object, but the former should not be confused with the latter. Arguably, for an agent to visually perceive or experience the location of an object, the representation of its spatial position must be transformed into either some egocentric frame of reference (centred on the agent's body) or some allocentric frame of reference (centred on some item belonging to the visual distal array).[11] Just as we now know more and more about the rules underlying the transformations of coordinate systems by means of which the spatial location of a stimulus can be represented, so we also need a computational framework for assessing the functional significance of the discovery of the anticipatory activity of LIP neurons that represent the location of a stimulus in retinocentric coordinates to the visual perception of distal objects whose location must be represented in different frames of reference for different purposes.

In my view, the best computational framework for assessing the functional significance for visual perception of such mechanisms as the anticipatory (or 'pre-emptive') activity of LIP neurons is provided by what computational theorists of action (such as Wolpert 1997 and Wolpert et al 1998) call "internal models of action". Now, computational

[11] In fact, there is much evidence from the neuropsychological examination of brain-lesioned human patients (eg DF) that representing the location of a graspable object in egocentric coordinates is necessary for grasping it, but is not sufficient for enabling the agent to be visually aware of its shape. This observation is crucial evidence in favour of the two-visual-systems model of human vision (see the previous section).

theorists of action distinguish two classes of internal models of action: forward models and inverse models. Roughly speaking, forward internal models of action take as input information about an agent's motor instruction and they compute the sensory consequences of the agent's intended motor act. Inverse internal models of action take as input a representation of the agent's goal (or sensory consequences of the intended action) and compute the motor command appropriate to achieve the agent's goal. In the rest of this section, only *forward* internal models will be relevant.

According to the forward-internal-models theory of voluntary (ie self-generated) action, the human motor system is richly endowed with predictions of the sensory consequences of the movements of one's own body (cf Blakemore et al 2000; Jeannerod 2003; Tsakiris and Haggard 2004). As illustrated by Wurtz and Sommer's (2004) and Duhamel et al's (1992) research on the oculomotor system, each time an agent's motor intention to act sends a motor command to some of the agent's muscles, it also sends a corollary discharge or efference copy (of the motor command) to some central monitor for planning, controlling, and checking the execution of the action. As I said, a forward model of action takes the efference copy of an impending motor instruction as input, and computes the sensory consequences of the intended motor act. Before executing eg the reaching-to-grasp movement, one can, from the efference copy of one's own motor command, predict that if and when one moves the right hand forward, grasps the glass with the right hand and lifts it above the table, one's visual experience will represent the glass as being above the table, not on it. As a result, when one indeed grasps the glass and lifts it up, unlike an observer, one is not surprised when one become visually aware that the glass no longer rests on the table, since this was a predicted consequence of one's action.

From an agent's point of view, what is the main difference between a sensory change that arises from one of his/her own self-generated movements and a sensory change that has been exogenously caused (ie not endogenously caused)? The main difference is that, unlike the latter, the former is *predictable* by the agent. According to the theory of forward internal models of action, what enables an agent to predict the sensory consequences of the impending motor act is precisely the existence of a corollary discharge or an efference copy of his/her motor instruction. Now sensory changes that are predictable by an agent because they arise from the agent's own self-generated movements should be less surprising than sensory changes that are not predictable for lack of an efference copy of the impending self-generated movement. Being less surprising, an agent's visual experience of a sensory change that arises as a result of the self-generated movement should be *attenuated*. By contrast, an agent's visual experience of a sensory change that is exogenously caused (and therefore not predictable from an efference copy of the agent's motor instruction) should be highlighted. At least, an agent could discard or minimise the visual processing of predictable sensory changes in his/her environment and pay more visual attention to exogenously caused (and therefore unpredictable) sensory changes.

Thus, the theory of internal forward models of action predicts that the more a sensory change is predictable by an agent, the more its experience will be attenuated and conversely the less predictable a sensory change, the more its experience will be highlighted. Empirical support for this prediction comes from two sources. On the one hand, Blakemore et al (1999, 2000) have showed that healthy human adults turn out to be unable to tickle themselves: the movements that produce an experience of ticklishness must be unpredictable. If the movements whose goal is to produce an experience of ticklishness are self-generated, then their sensory consequences (ie ticklishness) are predictable on the basis of corollary-discharge information. If so, then the experience of their sensory consequences is attenuated. On the other hand, Frith et al (2000) and Frith (2005) have argued that schizophrenic patients with either auditory

hallucination or delusions of control show abnormally strong awareness of the sensory consequences of their own actions and self-generated movements. If so, then we can give a pretty good sense to the idea that the anticipatory activity of LIP neurons are mechanisms of 'pre-emptive perception': one function of anticipatory mechanisms (based on corollary-discharge information) is to 'pre-empt' (ie to attenuate) the experience of predictable sensory changes (ie sensory changes that are predictable by an agent on the basis of corollary-discharge information about the impending self-generated movements).

But if so, then mechanisms of 'pre-emptive perception' do not support the enactive conception of visual experience at all. Remember: according to the enactive view, the main source of what philosophers call 'the phenomenal character' (or the phenomenology) of one's perceptual experiences is one's sensorimotor knowledge of the sensory consequences of one's bodily movements. On the enactive view, the content of one's visual awareness depends on one's ability to predict sensory changes that arise from one's own actions. To pre-empt, however, is not to enact. Indeed, contrary to the enactive view that one's sensorimotor knowledge of the sensory consequences of one's own bodily movements is the main source of the phenomenal content of one's perceptual experiences, the theory of forward internal models of action suggests that, because they are predictable, the sensory consequences of one's own actions should not give rise to perceptual experiences endowed with much phenomenal content (or phenomenology). The more a sensory event is predictable by an agent (as a result of corollary-discharge information about one's own intention to act), the less attention grabbing the sensory event and the more attenuated the experience of the event.

5 Conclusion
In this paper, I have discussed a view of visual perception and visual experience that has recently arisen as a result of a dissatisfaction with the traditional computational and representational paradigm: the so-called 'enactive' view. This view emphasises the constitutive contribution made by one's sensorimotor knowledge of the sensory consequences of one's own actions to the content of one's visual experiences. In a nutshell, I have made three points. First, I have pointed out some implausible metaphysical consequences of enactivism. Secondly, I have argued that enactivism is hard (if not impossible) to reconcile with all the evidence in favour of the two-visual-systems model of human vision. Finally, I have argued that, on the most plausible interpretation, mechanisms of so-called 'pre-emptive perception' cannot be enrolled by enactivists.

References
Aglioti S, DeSouza J F X, Goodale M A, 1995 "Size-contrast illusions deceive the eye but not the hand" *Current Biology* **5** 679 – 685
Blakemore S-J, Frith C D, Wolpert D M, 1999 "Spatio-temporal prediction modulates the perception of self-produced stimuli" *Journal of Cognitive Neuroscience* **11** 551 – 559
Blakemore S-J, Wolpert D, Frith C D, 2000 "Why can't you tickle yourself?" *NeuroReport* **11** R11 – R16
Block N, 2005a "Review of Alva Noë's *Action in Perception*" *Journal of Philosophy* **CII** 259 – 272
Block N, 2005b "Two neural correlates of consciousness" *Trends in Cognitive Sciences* **9** 46 – 52

Bodis-Wollner I, 2005 "Proposal for an international conference: Pre-emptive perception", personal communication

Clark A, Chalmers D J, 1998 "The extended mind" *Analysis* **58** 10 – 23

DeLoache J S, Uttah D H, Rosengren K S, 2004 "Scale errors offer evidence for a perception – action dissociation early in life" *Science* **304** 1027 – 1029

Dretske F, 2004 "Change blindness" *Philosophical Studies* **120** 1 – 18

Duhamel J-R, Colby C L, Goldberg M E, 1992 "The updating of the representation of visual space in parietal cortex by intended eye movements" *Science* **255** 90 – 92

Franz V H, Fahle M, Bülthoff H H, Gegenfurtner K R, 2001 "Effects of visual illusion on grasping" *Journal of Experimental Psychology: Human Perception and Performance* **27** 1124 – 1144

Franz V H, Gegenfurtner K R, Bülthoff H H, Fahle M, 2000 "Grasping visual illusions: no evidence for a dissociation between perception and action" *Psychological Science* **11** 20 – 25

Frith C D, 2005 "The self in action: lessons from delusions of control" *Consciousness and Cognition* **14** 752 – 770

Frith C D, Blakemore S-J, Wolpert D M, 2000 "Abnormalities in the awareness and control of action" *Philosophical Transactions of the Royal Society of London, B. Biological Sciences* **355** 1771 – 1788

Glover J, 2004 "What causes scale errors in children?" *Trends in Cognitive Sciences* **8** 440 – 442

Goodale M A, 1995 "The cortical organization of visual perception and visuomotor control", in *An Invitation to Cognitive Science, Visual Cognition* volume 2, Ed. D Osherson (Cambridge, MA: MIT Press)

Goodale M A, Milner A D, Jakobson I S, Carey D P, 1991 "A neurological dissociation between perceiving objects and grasping them" *Nature* **349** 154 – 156

Haffenden A M, Schiff K C, Goodale M A, 2001 "The dissociation between perception and action in the Ebbinghaus illusion: non-illusory effects of pictorial cues on grasp" *Current Biology* **11** 177 – 181

Hauser M D, 2003 "Knowing about knowing, dissociations between perception and action systems over evolution and during development" *Annals of the New York Academy of Sciences* **1** 1 – 25

Jacob P, 2005 "Grasping and perceiving objects", in *Cognition and the Brain, the Philosophy and Neuroscience Movement* Eds A Brook, K Akins (Cambridge: Cambridge University Press)

Jacob P, 2006 "Why visual experience is likely to resist being enacted" *Psyche* **12** 1 – 11; http://www.psyche.Cs.monach.edu.au/

Jacob P, Jeannerod M, 2003 *Ways of Seeing: The Scope and Limits of Visual Cognition* (Oxford: Oxford University Press)

James .., 2003 ....

Jeannerod M, 2003 "The mechanism of self-recognition in humans" *Behavioral Brain Research* **142** 1 – 15

Jeannerod M, Decety J, Michel F, 1994 "Impairment of grasping movements following bilateral posterior lesions" *Neuropsychologia* **32** 369 – 380

Jeannerod M, Jacob P, 2005 "Visual cognition: a new look at the two visual systems model" *Neuropsychologia* **43** 301 – 312

Lamme V, 2003 "Why visual attention and awareness are different" *Trends in Cognitive Sciences* **7** 12 – 18

Mahon B Z, Caramazza A, 2005 "The orchestration of the sensory – motor systems: clues from neuropsychology" *Cognitive Neuropsychology* **22** 480 – 494

Marr D, 1982 *Vision* (San Francisco, CA: Free Press)

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

Milner A D, Paulignan Y, Dijkerman H C, Michel F, Jeannerod M, 1999 "A paradoxical improvement of misreaching in optic ataxia: new evidence for two separate systems for visual localization" *Proceedings of the Royal Society of London* **266** 2225 – 2229

Milner A D, Perrett D I, Johnston R S, Benson P J, Jordan T R, Heeley D W, Bettucci D, Mortara F, Mutani R, Terazzi E, Davidson D L W, 1991 "Perception and action in 'visual form agnosia'" *Brain* **114** 405 – 428

Noë A, 2004 *Action in Perception* (Cambridge, MA: MIT Press)

O'Regan J K, 2000 "Change blindness", in *Encyclopedia of Cognitive Science* http://nivea.psycho.univ-paris5.fr/ECS/ECS-CB.html

O'Regan J K, Noë A, 2001 "A sensorimotor account of vision and visual consciousness" *Behavioral and Brain Sciences* **24** 939 – 1031

Rowlands M, 1999 *The Body in Mind* (Cambridge: Cambridge University Press)

Simons D J, Rensink R A, 2005 "Change blindness: past, present, and future" *Trends in Cognitive Sciences* **9** 16 – 20

Sommer M A, Wurtz R H, 2002 "A pathway in primate brain for internal monitoring of movements" *Science* **296** 1480 – 1482

Tsakiris M, Haggard P, 2004 "Experimenting with the acting self" *Cognitive Neuropsychology* **21** 1 – 19

Wolpert D M, 1997 "Computational approaches to motor control" *Trends in Cognitive Sciences* **1** 209 – 216

Wolpert D M, Miall R C, Kawato M, 1998 "Internal models in the cerebellum" *Trends in Cognitive Sciences* **2** 338 – 347

Wurtz R H, Sommer M A, 2004 "Identifying corollary discharges for movement in the primate brain" *Progress in Brain Research* **144** 47 – 60