

Is there a vertebrate mechanism of consciousness, and how would we recognize it if we saw it?

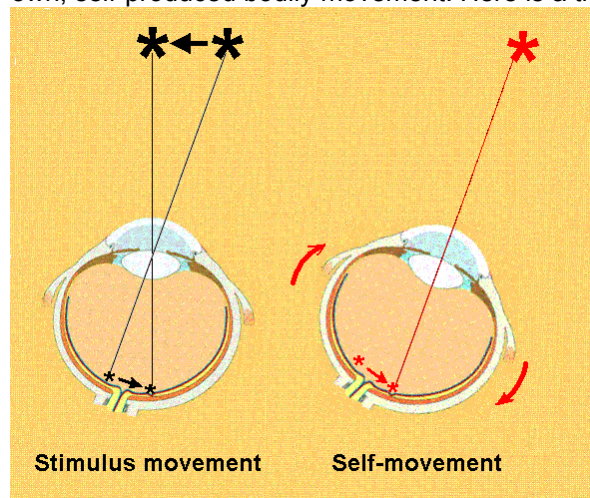
Bjorn Merker - ASSC11, 2007, Las Vegas

Transition from Irene Pepperberg:

You have just heard an elegant exposition of what might be called a functional or computational equivalence argument for the attribution of consciousness to a nonhuman species. Where there is comparable function, there may be a comparable mechanism. Thus, if in our case, consciousness is a feature of a certain mode of sophisticated cognition, it may be a feature of non-human instances of such a mode of cognition as well. For this to be so, it is not necessary that consciousness evolved *for* that mode of cognition, but perhaps for a more elementary function which in turn became a stepping stone to higher forms of cognition. That would account for the linkage between advanced cognition and consciousness while at the same time leaving wide open the possibility that species without these advanced capacities might be possessed of consciousness as well. Let me then outline for you a functional conception of consciousness that may fit this bill, and so might serve as a heuristic in the search for animal consciousness more generally.

One of many obstacles to clarity about consciousness is that its contents are so fundamentally and robustly useful that we take them for granted. One of those contents is the external world, which in plain view surrounds us on all sides... In fact, of course, what we see around us is a highly derived product of the brain's activity, and the reason the brain synthesizes such an external world is presumably part of its main business: to guide a physical body in getting about in a physical universe to fulfill the needs of that body. The fact that under ordinary circumstances this visible world bears no trace of the busy neural activity that continually creates it, is itself a major clue to the nature of consciousness: only the final polished products of the brain's handiwork end up in it, and in so doing every trace of the preliminaries that went into that handiwork are stripped and excluded from the synthetic product. That in turn means that to maintain such splendid isolation from the many and complex streams of neural activity that are its necessary precursors, consciousness must somehow be a localized, less than global, functional system. The reasons for this I think are intimately linked to the brain's predicament in being charged with guiding its body in the world by, as it were, remote control.

The brain's basic problem in that regard is of course that it has no direct access to either the physical universe or the physical body it must guide through it, because the brain is sitting enclosed in an escape-proof prison, the skull. Essentially it is, as it were, dark in there, and the only thing that surrounds the brain is biological tissues; it has no direct contact with anything else. For everything it must accomplish it must rely on the information it can glean indirectly, from the buzzing of sensory nerves, and on whatever control it can exert by making muscles contract by squirting them with acetylcholine. And each time it squirts, the resulting bodily movement is likely to displace one or another of the receptor arrays that activate the sensory nerves entering the brain, and so the activity in those nerves cannot even tell the brain for sure whether it reflects an event in the world or the brain's own, self-produced bodily movement. Here is a trivial example:



Notice that it is not possible to correctly interpret the optic nerve output - (which is identical in these two situations – stimulus movement and eye movement) - without having some means to distinguish self from other, a first hint that these matters may bear on the problem of consciousness. The example, of course, is only one small and elementary instance of a massive set of so called inverse problems the brain confronts in its confinement inside an opaque skull, problems that plague it no less on the motor control than on the sensory sides of its operations.



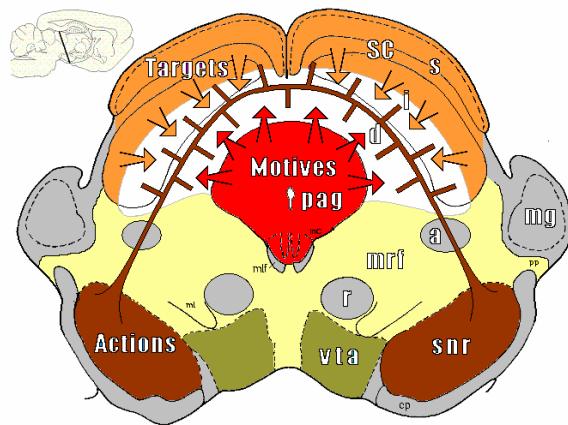
To illustrate the nature of so called inverse or ill posed or underconstrained problems, consider the predicament of this M C Escher character. I have updated this panorama from the days of the Lewis & Clark Expedition, featuring the main tributaries of the Missouri River, by placing polluters on each of the tributaries [explain by pointing them out]. This man's task is to try to figure out the sources and pattern of pollution upstream on the basis of measurements conducted in the Missouri itself, where he is sitting. Everything is of course strictly causal, but not in billiard ball fashion, so for any given pattern of measurements downstream there might be several – and indeed innumerable – patterns of pollution which in combination with relative tributary flow rates and other causal conditions generate identical mixture patterns downstream. Even determining how many tributaries (and polluters) there are on the basis of downstream measurements is a challenge which may or may not have a unique solution, depending upon circumstances. This is the inverse problem in a nutshell.

And like our M C Escher character contemplating this problem by way of an impossible figure, the brain's confinement inside the skull forces it to try to reconstruct what is going on "upstream" (i.e. out in the physical world) on the basis of the "downstream" effects induced by that world in the proximal receptors and nerves that are the brain's only source of information regarding that world. Even when our brain sends us scurrying off to a perceived potential target or source to inspect it at close quarters we are not going "upstream" to take direct measurements, but only adding new proximal signals, with new demands for interpretation. All the brain's measurements are made on the body surface, which itself is remote from the brain, and moreover multiply mobile with respect to the world and itself. In fact, brains employ whole specialized sensory systems, the proprioceptive as well as vestibular systems, to inform themselves of the disposition of the bodies they inhabit, and the state of their joints and muscles in particular.

And trying to reconstruct what is going on around it on the basis of its indirect measurements, the brain could do worse than to build itself a little model of what appears to be going on, and to use predictions from this model to refine its estimates and control parameters, that is, what engineers grappling with similar problems in the control of complex automated systems such as power-grids call a forward model. Formalisms from this domain have been explored in application to both sensory and motor aspects of neural function by Mitsuo Kawato, Daniel Wolpert and others. Some of the control problems the brain faces are hardly tractable without relying on solutions from this domain, and the brain presumably employs a number of neural forward models to bootstrap itself out of its multiple inverse problems. The most concise way of formulating what I am proposing regarding consciousness is this: The brain uses a centrally placed forward model for efficient control of gaze deployment, and the format in which this forward model operates amounts to a conscious mode of functioning.

Movements of the gaze, or more precisely of the orienting reflex in which gaze movements play a primary part, take the lead in all behaviors. This has been shown in exquisite detail for reaching and

manipulation by Roland Johansson, Randall Flanagan and others. [Act it out, and mention talks by Durgin, Kuhn, and comments by the magicians, regarding the leading role of the gaze]. This leading role of the gaze in behavior means that “the very next action” in real time is always a gaze movement. This gives the gaze a pivotal role in the brain’s control economy, and makes gaze control the brain’s principal output. But that means that the entirety of the brain’s operations – its best estimates of what is going on and what needs to be done – must be made available to the mechanism of gaze control, and this incessantly and continuously, in real time, throughout our waking hours in crisis as well as in leisure. As I have outlined in a recent *Behavioral and Brain Sciences* target article, savings are achievable in that incessant updating process by interfacing three crucial functions at the penultimate stage of gaze control, namely the functions of target selection, action selection, and motivational ranking. These sum up, as it were, the brain’s principal tasks, and these functions are in fact interfaced exactly where they are needed if their interaction is to control the gaze movements of orienting, namely in the intermediate and deep layers of the superior colliculus of vertebrates. Here, in rough outline, is a sketch of how:

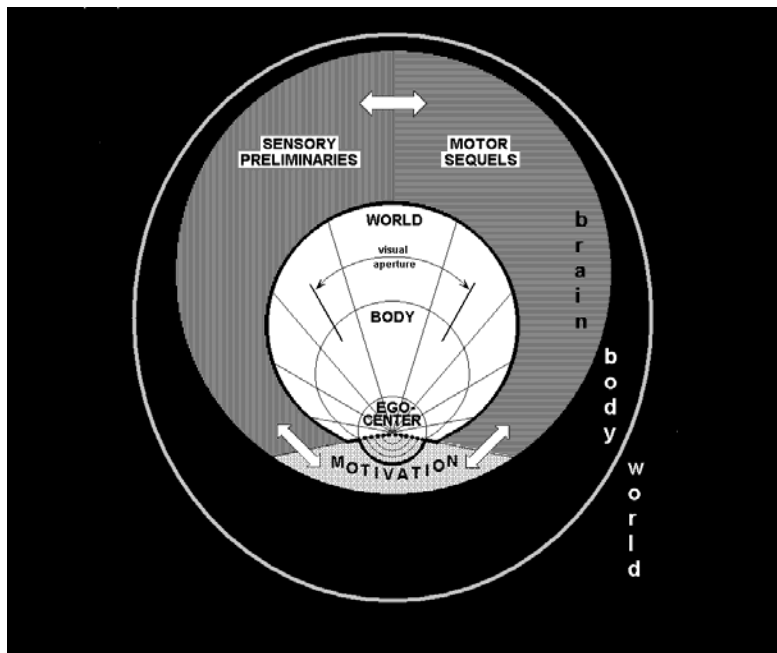


The selection triangle:

Target selection, Action selection, Motivational ranking

Selection Triangle. Comment: Here then [point to arch of intermediate-to-deep colliculus], is one place in the brain “where it all comes together”, and this “coming together” is not only in space, but in time as well, since conduction times along the short retino-collicular and the long retino-geniculo-cortico-collicular visual paths are matched, yielding coincident arrival times in the colliculus. So if you have been told that “there is no place in the brain where it all comes together” you have been deceived.

Now, the reason I think that this is relevant to mechanisms of consciousness is that the gaze movements of orienting, at their most basic and typical, consist of rotatory displacements of eyes in head and of head on trunk. That means that an efficient format for the neural control of the gaze is an arrangement of nested analog maps, held together in a joint spherical coordinate system that nests a body map inside a multimodal world map, with the coordinate origin lodged in the head representation of the body map. Such an arrangement would allow the crucial coordinate transformations between eyes, head, and world, needed to stabilize the world during movement, to be performed by simple remapping in two dimensions: azimuth and elevation. Now, such a nested arrangement of body and world around their shared coordinate system origin provides a striking match to the general outlines of our sensory consciousness, which features an implicit ego-center located inside the head representation of our body, from which we gaze out at our synthetic world through a cyclopean visual aperture. The leading role of the gaze in behavior means that this spatial orienting framework of ego-center, body and world in control of gaze movements must and will be kept continually informed of the brain’s best estimate of what to do next in light of current circumstances and needs. Here then is my most abstract sketch of the structure of what I have just outlined:



(POINT) This unshaded area, then, is what I called the “little model” the brain builds to help itself in its inverse predicament. And this part of the model, in its fully elaborated human implementation, is the external world we see around us. That is, what we see around us is not *this* (black world) but *this* (white world), though obviously there is a correspondence, a veridical relationship, between *some aspects* of this (black) and this (white). That, after all, is the whole point of incorporating what I called the “little model” in the brain’s control circuitry, though to us, occupying the position of ego-center inside that model, it is not “little” at all, but encompasses the full scope of our surroundings as far as the eye can see...

The ego-center-bodymap-worldmap system is thus no empty arrangement of spatial shells. It is filled, rather, not only with sensory content needed to target the gaze in space, but with the full play of motivational tensions that determine which targets are relevant and which not. Individuals are possessed of multiple needs, and the filling of those needs often involves incompatible behaviors, venues, or times of execution. Since the many needs cannot all be filled simultaneously they share and often compete for the time and other resources of that individual by dividing its available time between them. The tasks involved in this are therefore typically performed one at a time, sequentially. I think Michel Cabanac was the first to realize that consciousness is the medium of the hedonic common currency needed to make the trade-offs involved in that time-sharing. Thus, the various feelings we experience are the representatives in consciousness of our various motivational systems. They intrude on consciousness in the form of feelings to alert us to our needs, and to steer our orienting behavior accordingly.

What makes these considerations relevant to the question of animal consciousness is the fact that the logic of the analysis I just sketched is completely independent of the level of cognitive sophistication at which the contents of consciousness are elaborated in a given species. Whether an animal’s sensory world is a rich three-dimensional panorama like ours, or a mere two-dimensional screen on which environmental targets or threats appear as simple loci of motion, in either case savings are to be had in the economy of orienting over time by interfacing target selection, action selection and motivational ranking with the system of nested coordinate transformations needed for the efficient control of orienting responses.

In my figure illustrating the selection triangle I showed this interface implemented at the level of the midbrain, and given the conserved nature of brainstem circuitry throughout vertebrates, all vertebrates would accordingly be candidates for possession of a primary mode of phenomenal consciousness realized at this upper brainstem level, however threadbare its contents might appear compared to those of our own expanded mechanisms. But the same savings that supplied a selection pressure for the development of the brainstem interface in our own phylum are available to any mobile life form that finds its way to them, even from an utterly different starting point in terms of nervous system layout. For such forms, the vertebrate brain plan provides no anatomical guidance, and the question becomes how we would recognize a mechanism of consciousness implemented in the setting of a radically

different neuroanatomy. If my tentative analysis of the functional role of consciousness has any merit, it might provide at least *some* guidance even on excursions into such unfamiliar territory. Here are a couple of its inhabitants:



Slides 6 & 7 – jumping spider, octopus

From the perspective I have outlined, one might make a start on that question by examining the behavior and the neural control of a candidate species' orienting responses. That is, not its locomotor navigation, but its phasic orienting response to environmental events. Is that behavior under centralized control, and if so, do all of the animal's spatial senses converge on it in a premotor framework? Is the control circuitry for orienting supplied with compensatory input from an inertial-gravitational system analogous to the vestibular one, to stabilize the body-world relation during self-movement? Do available targets for orienting interact in real time with the animal's body deportment (spatial position, posture and movement) as well as with its prevailing need states in determining the target for the next orienting movement? If so, is that three-way interaction based on the brain's best estimate of each of the three interacting domains, so as to make the outcome of their interaction a best-case basis for triggering the next orienting movement? And finally, is the origin of the coordinate system for control of orienting the geometric pivot of a shared coordinate space within which the neural body-world relation is stabilized during self-movement?

Points summarized:

Is the control of the orienting response

- centralized?
- served by all spatial sensory modalities?
- supplied with inertial-gravitational input for stabilizing the world during movement?
- served by an interface between target selection, action selection, and motivational ranking?
- cast in nested format of ego-center, body map and multimodal world map?

Should the answer to all of these questions be yes, then I would move the species in question up among serious contenders for the status of conscious beings.

The emphasis on a centralized solution in the foregoing is motivated by the fact that we are interested in whether a given species is conscious or not, and not in whether it has solutions to the various control problems that are components of such a centralized solution. Every mobile animal confronts these problems, and solves them in one way or another, but far from all of them evolve a comprehensive, centralized solution to them. Insects provide numerous examples of successful piecemeal and sometimes surprisingly peripheral solutions to their control and prediction requirements.

Barbara Webb has reviewed some of these in Trends in Neurosciences of 2004. In decentralizing much of their neural control to segmental ganglia, their neural organization is quite unlike that of vertebrates. An ant will continue to live after the removal of its brain from its body, and in that condition it will exhibit many complex behaviors given suitable somatic stimulation. No vertebrate, of course, lives for more than seconds after the removal of its brain, because in vertebrates even vital functions are under central control. Or consider this: suitably positioned, a grasshopper will continue to feed actively on grass after its abdomen has been severed from its head end.

Such examples look bizarre from a vertebrate point of view, because they imply that these animals lack some form of integrated superintendence over the state of the animal as a whole. Indeed, their nervous systems perform complex and sophisticated functions on the basis of a decentralized mode of neural organization that lacks a central decision-making process. Circuitry implementing such modes of control have been traced in the insect nervous system by Jennifer Altman and Jenny Kien, and related control principles figure in the behavior-based approach to robotics pioneered by Rodney Brooks. Yet in the absence of a centralized nexus for decision-making, an animal will forego the savings achievable by competitively pitting the full range of influences bearing on its state as a whole against one another, and using the outcome of the competition as an input to the next orienting movement.

As part of a forward model for orienting, such a decision nexus amounts to a monitoring function for global superintendence of the affairs of the animal as a whole, and constitutes, I suggest, the hub of what we ourselves are, and call the “self” in its basic sense, stripped of all acquired characteristics. As the monitor function of a monitored domain, the self is an obligatory function of consciousness. Its presence defines the difference between unconscious processes and conscious ones: Only the contents of the latter are objects for a monitoring function – a self or subject. Self plus contents together constitute the whole of consciousness, in which self is the seeing but never seen part of consciousness, and the contents the seen but never seeing part, to use a very apt figure of speech from Buddhist psychology.

[That’s as far as I got in my talk, which was intended to end as follows:]

To round off on a personal note: When I first started actively pursuing these problems more than 35 years ago, I took it as a matter of course that what one is looking for in searching for a neural mechanism of consciousness is a tandem arrangement consisting of a monitoring function and a monitored domain. In neural terms that would mean looking for a miniature analog map – the monitor – reciprocally connected to a much larger and perhaps even multicomponent domain, the monitored content. Why? Because that is the most straightforward way to implement a monitor function by means of the brain’s favorite data format – topographic mapping. Moreover, a miniature analog map arrangement lends itself to swift and robust decision-making, implemented by local inhibitory connectivity in winner-take-all or similar fashion. As I searched the thickets of neuroanatomy to identify such an arrangement, I found that a number of systems and circuits turned up as promising candidates, rather than one unique one, as I had hoped. Here are a few of them:

Slide 9 – Funnel Anatomy [supplied in separate file, for clarity]. Explain by introducing Ann Graybiel’s terminology of “satellite system” for the parabigeminal nucleus (midbrain satellite of the colliculus), the thalamic reticular nucleus (diencephalic satellite of the thalamus), and claustrum (telencephalic satellite of cortex).

Slide 10 – Funnel diagram [supplied in separate file, for clarity]. Explain by pointing to the laminar differences in the cortical origin of the convergent projections constituting the several nested “funnels”, and the differences in the way the miniature analog map represents the different cortical areas. The claustrum by “grouping” all areas belonging to a given modality into single claustral zones; the thalamus by providing a nucleus per area, roughly speaking; and the colliculus by superposing areas in an exquisitely layered “map-stack” within a common, unified, premotor coordinate framework.

In the aggregate these miniature analog map and satellite arrangements define the thalamocortical complex plus its mesodiencephalic base. I can see no reason why each of these “funnels” might not make a unique contribution to the neural economy that issues in the full contents and dynamics of adult human consciousness. If so, there can be no satisfactory account of the brain’s mechanism of consciousness until the functional role of each of these analog map arrangements within the over-all

neural division of labor serving consciousness has been worked out. My own recent work has focussed on the significance of the mesodiencephalic base of this system. It holds special interest from the point of view of animal consciousness, since if my suggestions regarding it should turn out to have any merit, it is capable of sustaining a mode of primary phenomenal consciousness on its own. Its bare-bones functional principles, which I have tried to outline today, might not specify the very simplest mechanism of consciousness conceivable, but at least one for which we have some reason to believe that it would make a substantial contribution to the control economy of any species that found its way to implementing it. It is on that basis that I suggest it as a useful heuristic for approaching the problem of animal consciousness more generally.