Suppose that a team of neurosurgeons and bioengineers were able to remove your brain from your body, suspend it in a life-sustaining vat of liquid nutrients, and connect its neurons and nerve terminals by wires to a supercomputer that would stimulate it with electrical impulses exactly like those it normally receives when embodied. According to this brain-in-a-vat thought experiment, your envatted brain and your embodied brain would have subjectively indistinguishable mental lives. For all you know—so one argument goes—you could be such a brain in a vat right now.¹

Daniel Dennett calls this sort of philosophical thought experiment an “intuition pump” (1995). An intuition pump is designed to elicit certain intuitive convictions, but is not itself a proper argument: “Intuition pumps are fine if they’re used correctly, but they can also be misused. They’re not arguments, they’re stories. Instead of having a conclusion, they pump an intuition. They get you to say ‘Aha! Oh, I get it!’” (Dennett 1995, 182).

Philosophers have used the brain-in-a-vat story mainly to raise the problem of radical skepticism and to elicit various intuitions about meaning and knowledge (Putnam 1981). The basic intuition the story tries to pump is that the envatted brain, though fully conscious, has systematically false beliefs about the world, including itself. Some philosophers reject this intuition. They propose that the envatted brain’s beliefs are really about its artificial environment or that it has no real beliefs at all. According to these proposals, the mental lives of the two brains do not match, despite their being subjectively indistinguishable.

Dennett (1978) tells a classic variant of the brain-in-a-vat story, one in which he sees his own envatted brain and knows that it remotely controls his own body, but still cannot experience himself as located where his brain is located. Here the thought experiment serves to raise questions about the locus of the self in the physical world.
Underlying these varied uses and rival assessments lies a fundamental, shared intuition—that a suitably working human brain is not only necessary, but also sufficient all on its own for the instantiation or realization of our subjective mental life. Given our knowledge that the physical processes crucial for human mentality occur in the human brain, it seems imaginable that these processes could occur in the absence of the rest of the body, as long as the right causal supports were provided, and that such disembodiment would make no difference to our subjective experience. This idea is the deeper “Aha! Oh, I get it!” intuition the brain-in-a-vat story pumps.

As Dennett notes, philosophers often fail to set up their intuition pumps properly by failing to think carefully about the requirements and implications of their imagined scenarios. Brain-in-a-vat stories typify this shortcoming. Philosophers help themselves to this scenario and the basic intuition that it is supposed to pump without thinking about what the scenario actually demands of our imagination when we try to spell out the story in sufficient detail. In this way, they make substantive empirical assumptions about the biological requirements for consciousness that may well be false, and they ignore the difficult conceptual problem of how to distinguish within those biological requirements between what contributes only causally to the production of subjective experience and what constitutes or instantiates or realizes subjective experience.

We propose to take Dennett’s advice to heart and think carefully about the details of this thought experiment. Given our knowledge of the brain, what do we need to specify in order to imagine properly a brain in a vat? In addressing this question, we intend to put the brain-in-a-vat thought experiment to a new use, namely, to address the biology of consciousness and to develop some new considerations in support of the enactive approach in cognitive science (Thompson 2007).

13.1 The Argument in a Nutshell

When theorists invoke the notion of a brain in a vat, they invariably take a unidirectional control perspective and view the brain as a kind of reflexive machine whose activity is externally controllable. Yet numerous neurobiological considerations count against this viewpoint and indicate that the brain needs to be seen as a complex and self-organizing dynamical system that is tightly coupled to the body at multiple levels. The following points in particular deserve mention:
Brain activity is largely generated endogenously and spontaneously.
- Brain activity requires massive resources and regulatory processes from the rest of the body.
- Brain activity plays crucial roles in life-regulation processes of the entire organism and these processes necessitate the maintenance of viable sensorimotor coupling with the world.
- Thus the neurally enlivened organism meets the world on its own, endogenously generated sensorimotor terms.

Given these points, we propose the following null hypothesis for the brain-in-a-vat thought experiment: Any truly functional “vat” would need to be a surrogate body subject to control by the brain. By “body” we mean a self-regulating system comprising its own internal, homeodynamic processes and capable of sensorimotor coupling with the outside world. In short, the so-called vat would be no vat at all, but rather some kind of autonomous embodied agent.

This supposition has an important implication. It implies that our default assumption should be that the biological requirements for subjective experience are not particular brain regions or areas as such, but rather some crucial set of integrated neural-somatic systems capable of autonomous functioning. This assumption is one of the core working assumptions of the enactive approach (Thompson 2007).

13.2 Enactive versus Neurocentric Views of Consciousness

Before looking at the supporting evidence for the previous argument, we need to introduce some important concepts and distinctions. We can begin with the following enactive proposal about the brain basis of consciousness: “We conjecture that consciousness depends crucially on the manner in which brain dynamics are embedded in the somatic and environmental context of the animal’s life, and therefore that there may be no such thing as a minimal internal neural correlate whose intrinsic properties are sufficient to produce conscious experience” (Thompson and Varela 2001, 425). According to this proposal, the processes crucial for consciousness are not confined to the brain, but include the body embedded in the environment. For example, somatic life-regulation processes contribute to affect and sense of self (Damasio 1999), and dynamic sensorimotor activity contributes to the qualitative content of perceptual experience (Hurley and Noë 2003).

Ned Block (2005a) has recently argued that this sort of proposal fails to distinguish clearly between causation and constitution, that is, between
what causally contributes to consciousness and what neurobiologically constitutes consciousness. In the orthodox view, although conscious experience causally depends on the body and environment, it is directly determined by brain activity. This view can be given either a neuroscientific or philosophical formulation. According to the neuroscience version, some specific neural system or set of neural processes is the minimal biological substrate for conscious experience. According to the philosophy version, some specific set of neural processes is the minimal sufficient condition or minimal supervenience base or minimal realizing system for conscious experience.

A proper statement of this orthodox view, however, requires some refinements. We need to distinguish between the core neural realization and the total neural realization of consciousness or a given conscious state (Block 2005b; Chalmers 2000). In general, the core realization of a property or a capacity suffices for that property or capacity only when placed in the context of a larger system that constitutes the total realization (Shoemaker 1981; Wilson 2001). Block proposes that “the core NCC [neural correlate of consciousness] is the part of the total NCC that distinguishes one conscious state from another—the rest of the total NCC being considered as the enabling conditions for the conscious experience” (Block 2005b, 47).

Yet even this formulation remains incomplete. In general, the total realization of a property or a capacity suffices for that property or capacity only given the appropriate background conditions (Wilson 2001). So the total neural realization suffices for consciousness only given certain background conditions, which in the normal case include nonneural parts of the body and the environment.

We can now give a fuller statement of the orthodox and neurocentric view of consciousness. Given the appropriate background conditions (e.g., in the body and the environment), the total neural realizer suffices for consciousness all by itself; the core neural realizer suffices to determine a given conscious state (as specified by its content) and thus to distinguish one conscious state from another.

When we spell out the neurocentric view in this way, we bring to light a number of important problems that have been largely ignored by philosophical discussions that rely on the causal-versus-constitutive distinction for the brain basis of consciousness. Here is a list of these problems:

• In the case at hand, it is not clear how to draw the causal/constitutive distinction. What are the criteria for determining what is causal and what is constitutive in the neurobiology of consciousness?
The same question can be raised about the notions of core realization, total realization, and background conditions. What are the criteria for drawing these distinctions and applying them to the neurobiology of consciousness?

A given core realizer and/or constitutive supervenience base are usually identified by appealing to what plays the most salient causal role with regard to the instantiation of some property. In the case of the brain basis of consciousness, however, what plays the most salient causal role in any given case is far from clear (see the next point).

The question “What plays the most salient causal role?” typically cannot be answered for complex (nonlinear) systems, such as the brain, by pointing to the behavior of individual elements independent of the context of all the other state variables of the system (Cosmelli, Lachaux, and Thompson 2007; Wagner 1999). In dense nonlinear systems in which all state variables interact with each other, any change in an individual variable becomes inseparable from the state of the entire system. In such cases, the distinction between regular causes (regularities in the system’s behavior) and singular causes (unique nonrepeatable events that change the system’s behavior) becomes meaningless (Wagner 1999), and there is arguably no core realizer for a given property or behavior less than the system itself.

Finally, causal salience is an interest-relative and context-sensitive notion. Therefore, we seem to have no independent grip on constitution or realization (as metaphysical notions) apart from particular explanatory contexts.

These considerations suggest the following thoughts. At present, we have no clear way to draw the line between what is constitutive and what is causal in the biology of consciousness. To draw this line, we would need to have a far more developed understanding of the brain as a complex system and how its activity as a complex system is related to the body and environment. In the absence of this knowledge, we cannot assume that the brain suffices to realize consciousness all on its own apart from the body, or that particular neural systems suffice to realize one or another conscious state independent of the rest of the brain and the body.

How, then, might we proceed in the face of these problems? One way would be to ask what we could remove on the bodily side from a normal embodied brain while still preserving consciousness. Indeed, the brain-in-a-vat thought experiment proceeds exactly this way, by assuming that we could remove the body entirely, as long as everything else in the brain were held constant. In this way, the body can be shown to be inessential for the realization of consciousness, that is, as merely causally supportive.
or enabling, but not constitutive. Yet what if it were not possible to hold everything in the brain constant in the absence of the body? If certain brain processes simply could not be realized in the absence of the body, and these brain processes included those crucial for consciousness, then we would have reason to believe that the body is not merely causally enabling for consciousness, but also constitutive. The argument of this paper is that the brain-in-vat thought experiment, when spelled out with the requisite detail, suggests precisely this result.

13.3 A Close Look at the Brain in a Vat

We now need to examine in some detail the supporting evidence for this argument. In particular, we need to consider: (1) the demands of keeping the brain alive and up and running, (2) the spontaneous and endogenous activity of the brain in relation to the body, and (3) what it takes to mimic precisely the stimulation the nervous system normally receives from the environment.

13.3.1 Keeping the Brain Up and Running

Before getting to the point where we can stimulate the envatted brain or nervous system in a way that duplicates the stimulation it normally receives from the body and environment, we need to keep it alive and functioning. This already is no mean feat.

First, we need some protective apparatus for the brain. This apparatus serves to replace the skull (and spine, if we choose to keep the spinal cord). To ensure the brain’s flotation, the protective device must be filled with a liquid analogous to the cerebrospinal fluid (CSF). This liquid needs to be able to remove waste products of neuronal metabolism and therefore needs to be continually recycled (Brown et al. 2004; Davson and Segal 1971; Segal 1993). One way to achieve such recycling would be to couple the protective fluid to the second thing we need—a circulatory system.

Almost everyone has experienced the intense dizziness that accompanies standing up fast and the resultant cognitive impairment. The unimpeded supply of blood to every part of the brain is critical for its functioning and by no means a trivial physiological accomplishment. To envat the brain, we must provide an adequate blood supply (or alternatively, a fluid with similar biochemical properties). For this task, we could probably choose to keep the vascular system in place as a delivery structure. Alternatively, in the true spirit of the thought experiment, we can imagine replacing the entire cerebral vasculature with some synthetic device that
shows similar properties of selective permeability and local and systemic responsiveness to the ongoing demands of the brain-to-be-maintained. Such local and systemic responsiveness is absolutely crucial. Without it, there would be no way to compensate for even minimal departures from homeostasis due to neuronal activity, with fatal consequences for our experiment.

As early as 1890, Roy and Sherrington proposed that there should exist “an automatic mechanism by which the blood supply of any part of the cerebral tissue is varied in accordance with the activity of the chemical changes which underlie the functional activation of that part” (1890, 105). The coupling of blood flow and neuronal activity is a well established and basic physiological fact known as functional hyperemia (Hyder, Shulman, and Rothman 1999; Raichle and Stone 1971; Shulman, Hyde, and Rothman 2002). Indeed, many of the results informing contemporary hypotheses about the relation between brain and mind come from measurements of neuronal activity (fMRI, PET) that rely on different aspects of this coupling (Logothetis and Pfeuffer 2004). Although the actual mechanisms underlying this tight coupling are not fully understood, it appears that a variety of processes participate in the regulation of local blood flow, including direct neuronal release of vasoactive metabolic factors (such as H\textsuperscript{+}, K\textsuperscript{+}, lactate, adenosine, glutamate-induced neuronal production of nitric oxide, and several neurotransmitters; see Kandel, Schwartz, and Jessell 2000; Krimer et al. 1998; Paspalas and Papadopoulos 1998; Yang et al. 2000), astrocyte-mediated K\textsuperscript{+} siphoning from active synaptic regions to the local microvasculature, and Ca\textsuperscript{2+} dependent release of vasoactive molecules through the astrocyte perivascular end-feet (Anderson and Nedergaard 2003; Zonta et al. 2003). Our sustaining system must therefore be capable of responding locally to these (and probably other) factors in a highly specific and efficient way in order to sustain the local needs arising from ongoing neuronal activity. It is not difficult to see that any such synthetic apparatus would probably be as sophisticated as an actual vascular system in both its structural features and functional capacities.

Suppose we have succeeded in setting up such an immensely complex system. It would then be necessary to move the fluid through the delivery structure. Here some kind of pump is needed, as well as some minimal and highly selective recycling system for replenishing the fluid’s necessary components, including oxygen, glucose, and the numerous soluble ions, proteins, and other biomolecules that account for the fluid’s osmotic, nutrient, and regulatory properties.
This pump and recycling system needs to be responsive to the brain’s actual demands. To achieve this goal, some level of the brain’s activity needs to be coupled to the functioning of the circulatory system. Such coupling would ensure the local availability of the soluble factors provided by the circulatory system and would keep the concentration of the circulating molecules and ions within a physiological range despite continuous demands from the neuronal tissue.³

To meet these needs, the brain normally relies on a series of mechanisms involving multiple regulatory loops (in addition to those controlling the mechanical circulation of the blood). For example, the main neuroendocrine regulatory loop responsible for the control and allocation of glucose—probably one the most important parameters for brain function (Kandel, Schwartz, and Jessell 2000; Peters et al. 2002; Peters et al. 2004)—is the limbic-hypothalamic-pituitary-adrenal (LHPA) axis. Under normal conditions, the brain controls the allocation of glucose through at least two mechanisms. On the one hand, a tight coupling exists between neuronal activity and glucose uptake from the blood through the astrocyte end-feet glucose transporter GLUT1. This local and rapid on-demand mechanism depends on synaptic glutamate release and an adequate electrochemical sodium gradient across the astrocytic membrane, and therefore already represents an important level of coupling between energy availability and ion balance. On the other hand, the brain can regulate the level of glucose in the blood through the LHPA system, whereby it controls the release from adrenal cortex of cortisol (which acts as a feedback signal to control the activity of the LHPA system through corticosteroid brain receptors) and adrenalin release from the adrenal medulla (through sympathetic innervations). The release of these hormones—along with the inhibition of insulin release from the pancreas and suppression of muscular and adipose tissue glucose uptake—results in the rise of glucose concentration in the blood in a manner that directly depends on the actual workings of the brain (and probably the body). The multiplicity and complexity of analogous regulatory loops involving organs outside the brain, including such factors critical to neuronal function as electrolyte balance (Hebert, Brown, and Harris 1997; Simard and Nedergaard 2004; Yano, Brown, and Chattopadhyay 2004) and water homeostasis (Amiry-Moghaddam and Ottersen 2003; Grubb, Raichle, and Eichling 1977), can be seen by reviewing any standard physiology textbook.

Let us summarize the discussion up to this point. However simplified the life-sustaining system we produce for a brain in a vat, this system must involve at least the capacity to keep up with the energetic, ionic, osmotic,
and recycling needs of the brain. It will therefore include some kind of circulatory system, plus the necessary pumps, oxygenating devices, and additional subsystems for ensuring the maintenance of physiological levels in the circulating fluid. These points are obvious. The following point, however, is not so obvious: what the brain requires at any given instant depends on its own ongoing, moment-to-moment activity. Therefore, the life-sustaining system must not only be supportive of this activity, but also locally and systemically receptive and responsive to it at any given instant, independent of any external evaluation of the brain’s needs. Consequently, to keep the brain alive and functioning, this responsive system will most likely need to be energetically open, and self-maintaining in a highly selective manner. In other words, it will need to have some kind of autonomy. This system is starting to look less like a vat and more like a body.

13.3.2 Life, Homeodynamics, and the Body-Coupled Brain

In trying to fill in some of the design specifications for a system capable of keeping an envatted brain up and running, we began by taking an external and unidirectional control perspective. From this perspective, the issue is how to control the brain from outside so that it remains alive and functioning. Yet once we take into account the brain’s endogenous workings, it becomes obvious that our life-sustaining system must be intimately coupled to the nervous system’s labile activity at almost every level of this system’s construction and operation. This fundamental requirement necessitates a radical shift in how we think about our vat. Whatever life-sustaining system we construct, the functioning of its every part, as well as its overall coordinated activity, must be kept within a certain range by the nervous system itself in order for the brain to work properly. Hence the external and unidirectional control perspective is not generally valid. Instead, our life-sustaining system and the brain need to be seen as reciprocally coupled and mutually regulating systems.

The tight coupling between brain and body lies at the heart of the maintenance of organismic unity (Damasio 1999; Shewmon 2001; Swanson 2002). On the one hand, the nervous system tightly couples to the functioning of the body through numerous regulatory loops; on the other hand, the body’s proper functioning ensures the brain’s persistence as a functional subsystem. The nervous system’s basic role is to ensure the maintenance of a homeodynamic regime. The nervous system evolved to coordinate movement—probably one of the most challenging threats to homeostasis—by systematically coupling motor and sensory surfaces while providing a stable, internal biochemical milieu (Swanson 2000, 2002). In
constructing our vat, we need to keep in mind this crucial fact of neurally mediated organismic integration, because it provides the basic reference point for understanding the significance of neuronal activation overall. According to a number of authors, it also provides the basic underpinnings for subjectivity or the phenomenal sense of self (Craig 2002; Damasio 1998, 1999; Panksepp 1998a, b; Parvizi and Damasio 2001; Saper 2002).

From da Vinci’s pithed frog to current studies on the tight relationship between damage to midbrain structures and comatose or persistent vegetative states, the importance of the nervous system for keeping the organism alive, awake, and behaving adaptively has been amply demonstrated (Blessing 1997). Yet specific proposals about how consciousness is related to bodily life-regulation have only recently appeared (Damasio 1999; Panksepp 1998a, b; Parvizi and Damasio 2001). According to these proposals, the physiological constitution of a stable yet dynamic “core self” acts as an essential organizing principle for consciousness and derives from the nervous system’s capacity to monitor and ensure the body’s integrity. If our envatted brain is to have a subjective sense of self comparable to that of an embodied brain, then we need somehow to preserve this core self for the brain in a vat. To appreciate the complexity of this requirement, it is worth mentioning a few details from these proposals about the neural constitution of the core self.

According to Damasio (1998, 1999), the nervous system provides a stable ongoing map of the body by continually tracing the state of the body through a series of core neural structures. This neural map constitutes a “proto-self” that provides a reference point for cognitive and conscious capacities, thereby anchoring these capacities in a fundamental life-preservation cycle (Parvizi and Damasio 2001). The relevant core neural structures comprise several levels of the neuraxis, including brainstem nuclei of bodily regulation, hypothalamus and basal forebrain, and insular and somatosensory cortices, including medial parietal areas. In this framework, interoception provides the organism with continuous updated information about the internal state of the entire body, not just the viscera (Craig 2002, 2003; Saper 2002). Signals converging onto core neural structures (mainly at the level of the brainstem) from proprioceptive, vestibular, visceral, and other internal sources, combined with corresponding efferent regulatory processes that keep these parameters within a tight domain of possible values, establish internal dynamical regularities that ensure the organism’s viability through changing internal and external conditions. To support “mental processes and behaviors conducive to further homeostatic regulation” (Parvizi and Damasio 2001, 151), global bodily signals
need to be integrated with the state of activation of the cortex. This integration occurs through brainstem nuclei providing a complex network of modulatory effects on cortical activity, while rostral structures (such as the amygdala, cingulate gyrus, insula, and prefrontal cortex) provide descending influences on these brain-stem structures.

The importance of the basic self-preserving and self-monitoring organization of the nervous system within the body is likewise a central theme in Panksepp’s work on affective neuroscience (1998a, b). According to Panksepp, a specific region in the midbrain, the periaqueductal gray (PAG), qualifies as a massive convergence zone where emotional and attentional circuits coming from rostral regions in the forebrain interact not only with sensory and vestibular signals converging from the adjacent colliculi and deep tectal areas, but also with motor maps present in the deep layers of the superior colliculi (SC) as well as motor signals from the mesencephalic locomotor region (Panksepp 1998b). Panksepp proposes that primary consciousness is more closely linked to internal motor processes than exteroceptive sensory processes. The primal motor map in the SC maintains more stable motor coordinates than do the corresponding sensory maps and thus provides a secure self-referential set of internal motor coordinates upon which various sensory and higher perceptual processes can operate (Panksepp 1998b). PAG constitutes the core of the visceral-hypothalamic-limbic axis responsible for the primitive self-centered emotional and motivational systems that interact with the cognitively oriented core of the somatic-thalamic-neocortical axis. Thus, in Panksepp’s view, PAG serves as the substrate for a primal affective and sensorimotor sense of self.

Although Damasio and Panksepp differ on various specifics, they converge on certain fundamental points. First, life-regulation processes involving neural mappings of the body constitute a core self that grounds both neural activity overall and neural activity relevant to consciousness in particular. Second, primary consciousness includes an invariant basal awareness that remains constant across changing sensory contents. Third, this basal awareness is structured by an affective sense of phenomenal selfhood and thus constitutes a minimal form of subjectivity. Finally, primary consciousness or subjectivity needs to be seen as a large-scale feature of the homeodynamic life-regulation processes effected by the nervous system.

One more type of evidence for the notion of a vigilant and homeodynamically dedicated brain bears mention here. This evidence comes from functional magnetic resonance imaging (fMRI) studies. By analyzing a series of studies showing systematic task-independent decrease in
activation in certain brain areas, Gusnard and Raichle (2001) uncovered a set of cortical regions that appear to be continuously active during the resting state and whose activity decreases only upon goal-directed behavior. These cortical regions fall into four main groups: (i) posterior medial cortices, including medial parietal regions (these regions form part of the proto-self for Damasio); (ii) inferior lateral parietal cortices that also show significant activation when recovering from anesthesia; (iii) ventral medial prefrontal areas, which interestingly receive convergence of internal bodily information and external sensory information through the orbital regions, and have strong connections to limbic structures, amygdala, ventral striatum, hypothalamus, PAG, and other brainstem nuclei; and (iv) dorsal medial prefrontal cortex, which is also active during self-directed behavior, such as monitoring one’s own mental state (Gusnard and Raichle 2001). These results are consistent with the hypothesis that the brain, during resting conditions, is in a state of active bodily self-monitoring.

Furthermore, Raichle discusses another relevant issue in this context from a cost-analysis perspective (Raichle 2006; Raichle and Mintun 2006). Given that the brain needs no more than 1 percent of its total “energy budget” to deal with environmental demands, maintaining endogenous activity within viable limits is probably the most relevant task for the brain. In our view, this point suggests that self-sustaining ongoing activity, which is crucially coupled to the functioning of the body, holds the highest level in the control of brain functioning.

Let us return once again to our brain in a vat. If the previously mentioned proposals and hypotheses are sound, and if we were able to set up a life-sustaining system that also enabled the brain to maintain these self-related homeodynamic regimes, then we would have reason to believe that some kind of phenomenal subjectivity had been realized or instantiated by the envatted brain. This instantiation of subjectivity would depend on the integrity of the regulatory loops both within the brain and between the brain and its supporting vat-system, for these loops are what ensure the existence of the self-sustaining domain of physiological activity crucial for subjectivity.

From a neurocentric and unidirectional control perspective, it would seem that the brain is the superordinate controller of these regulatory loops. But this perspective is one-sided. It overlooks basic physiology, which tells us that the brain’s functioning is also subordinate to the maintenance of bodily homeostasis. As we have seen, the nervous system’s activity is inextricably coupled to the body and subordinate to the integrity of regulatory processes that extend throughout the body. Thus brain and
body are simultaneously both subordinate and superordinate in relation to each other. Put another way, neither one is intrinsically subordinate or superordinate; rather, they are reciprocally coupled and mutually regulating.

The point we wish to stress now is that this sort of dense reciprocal coupling between neuronal and extraneuronal systems must be in place in order for our envatted brain to instantiate or realize the neural processes crucial for phenomenal selfhood or subjectivity. Hence the total realization base for the subjectivity of the envatted brain corresponds to the system constituted by the coupling of these neuronal and extraneuronal subsystems. In other words, the total realizer sufficing for subjectivity is the brain-plus-vat and not the brain alone.

What about the core realizer for subjectivity? Is it purely neural? It is difficult to say. If we could turn subjectivity on and off by affecting neuronal activation alone while leaving everything extraneuronal unchanged, then we would most likely conclude we had found the core neural realizer for subjectivity. Of course, unless our brain in a vat could somehow report its states to us, we would have no way of knowing whether we were turning subjectivity on and off. Philosophers are familiar with this sort of problem; it is a variant on the problem of other minds. It is not this problem, however, we wish to emphasize, but rather the following one. Given the dense reciprocal coupling between neuronal and extraneuronal systems, there can be no neural change without a cascade of changes in many extraneuronal parameters. Turning subjectivity on and off would entail systematic alteration of these extraneuronal parameters just as much as systematic alteration of the neuronal ones. As we have seen, any change in neuronal activation implies a departure from homeostasis that demands immediate physiological compensation and this compensation must itself be regulated by the nervous system. To use dynamical systems language, neuronal and extraneuronal state variables are so densely coupled as to be nonseparable. From this perspective, the core realizer for subjectivity looks to be nothing less than some crucial set of densely coupled neuronal and extraneuronal processes. If this is right, then there may be no such thing as a purely neural core realizer for subjectivity.

13.3.3 Mimicking Environmental Stimulation
We still need to consider what it would take to produce specific conscious states, distinguished by their sensory contents, in the envatted brain. Here the minimal requirement is to deliver stimulation to the neuronal
terminals that duplicates or matches precisely the stimulation the brain normally receives from the environment.

The first point to be stressed is that such stimulation would have to be delivered without disrupting the life-sustaining system already established. This point is crucial. Adequate stimulating devices need to be constructed so that they can be integrated seamlessly into the vat. The complexity of such devices cannot be underestimated. Imagine an artificial device capable of stimulating every fiber of the optic nerve in perfect correlation with the light pattern of the scene to be recreated, guaranteeing all the dynamic receptive field relations found originally among retinal cells, maintaining perfect synchrony with the exploratory motor efference of the brain as it scans through the virtual image, and updating its activity so as to match precisely the sensory reafference.

Our artificial stimulating devices must therefore meet two basic requirements. On the one hand, the stimulation delivered to the neuronal terminals must mimic that obtained by the embodied nervous system. On the other hand, the devices must not disrupt the overall homeodynamic domain of activity crucial for life-regulation and subjectivity. These requirements suggest that our artificial stimulating devices must themselves be subject to tight regulation from the nervous system through artificial sensorimotor loops.

It is worth considering in this connection some examples of the important role that peripheral, nonneuronal processes play in the generation of neural activity. Consider first the development of spinal reflex circuits (Schouenborg 2003, 2004). Here it is crucial that the sensorimotor circuit be finely matched to the periphery for functional adequacy. Sensory feedback from spontaneous muscle twitches (occurring during sleep and analogous to human fetal movements) is critical for adapting the connections in spinal reflex modules to body anatomy (Petersson et al. 2003) and for determining the somatotopic functional organization of the somatosensory cortex (Khazipov et al. 2004). As Schouenborg remarks, “It is not the afferent input per se that is important . . . but rather the sensory feedback resulting from activity in the sensorimotor system” (2004, 694).

Convergent work on the development of the auditory cortex also points to the crucial role of peripheral structures in adapted neural activity. Mrsic-Flogel and collaborators used a “virtual acoustic space” to enable infant ferrets to hear through virtual ears of mature animals. Their results showed how changes in spatial coding during development of the auditory cortex seemed to be entirely due to changes in peripheral nonneuronal sensory structures (Mrsic-Flogel et al. 2001; Mrsic-Flogel, Schnupp, and King 2003;
see also Grubb and Thompson 2004). This finding reinforces the point that the choice of which peripheral structures to use to stimulate the envatted brain is not trivial.

Recent work on a realistic model of the neuromuscular system responsible for feeding behavior in the mollusk *Aplysia* also reveals the importance of the tight coupling between central neuronal systems and peripheral nonneuronal ones (Brezina, Orekhova, and Weiss 2000, 2003a, b; Brezina, Horn, and Weiss 2005). Among other things, this work addressed the following question: given that central (neural) motor commands show stochastic behavior, whereas the periphery (the complex network of muscle and modulatory neurotransmitters and neuropeptides) presents a slow, history-dependent dynamics, to what extent is the peripheral system under the control of the nervous system? Brezina and collaborators show that the periphery works to a certain extent in a semi-autonomous manner (Brezina, Horn, and Weiss 2005). The nervous system does not control the peripheral musculature in a hierarchical master/slave fashion; rather, optimal performance emerges only from the collective behavior of the interacting neuromuscular system (central and peripheral) in a given environment. These authors suggest that the peripheral network is responsible for part of the predictive and control functions of the neuronal tissue. In their words: “In vertebrates as well as invertebrates, the structural and dynamical complexity of the periphery can be as large as that of the central nervous system, so that, seen more abstractly, *the computational capability of the periphery rivals that of the nervous system that is attempting to control it*” (Brezina, Horn, and Weiss 2005, 1523; our emphasis).

Similar co-dependence of functional outcome can be observed at the level of neuronal networks themselves. Network activity is determined both by the intrinsic properties of the network and the modulatory environment, mainly through the modulation of synaptic behavior (Marder 1998; Marder and Thirumalai 2002). Therefore, together with neuronal firing, complex modulatory interactions between central neuronal cells and peripheral nonneuronal elements determine the nervous system’s response.

These examples are intended to stress the immense complexity of the neural and extraneural interactions that ultimately determine brain activity in the living organism. The list of functional systems dependent on brain-body coupling to provide the organism with coherent perception of the world also includes the entire interoceptive, autonomic system (Craig 2002, 2003; Saper, 2002), vestibular-autonomic regulation (Balaban and
Porter 1998; Yates and Miller, 1998), balance and somatic graviception relying on hydrostatic properties of blood pressure and inertial mass of abdominal viscera (Mittelstaedt 1996, 1997; Vaitl et al. 2002), as well interaction between the senses occurring at both central and peripheral levels (Howard 1997).

Let us return to our brain in a vat. The foregoing kinds of complex dependencies of neural activity on peripheral, extraneural systems must somehow be established for our envatted brain in order to mimic precisely peripheral stimulation as well as the way the embodied brain responds to such stimulation. Given the computational complexity involved, it is hard to imagine how to accomplish this feat simply by stimulating the neuronal terminals with electrical impulses generated by a supercomputer (Dennett 1991). Rather, it seems that we must equip the brain with real sensorimotor systems. Furthermore, as we suggested earlier, the brain must be able regulate these peripheral systems. Thus, at any given moment, the state of the peripheral systems will depend on the brain’s endogenous dynamics, which always shapes the sensory inflow (Engel, Fries, and Singer 2001; Varela et al. 2001); the state of the central systems will depend on how the peripheral systems are operating and what they have provided. Once all these structural and dynamical features are added to our already self-maintaining and energetically open vat, however, our so-called envatted brain looks a lot less like a brain in a vat and much more like an autonomous sensorimotor agent.

13.4 An Evo-Devo Digression

Before we present the results of our reflections on the brain-in-vat thought experiment, it is worth reminding ourselves of some basic facts about the evolutionary and developmental biology of the nervous system.

From an evolutionary perspective, brain and body are co-evolved structures that match one another’s properties through a history of adaptive phylogenetic changes in different species (Aboitiz 1990, 1996; Chiel and Beer 1997; Funes and Pollack 1998). This fact already suggests that considering the brain as some kind of internal director of the organism uniquely responsible for its cognitive capacities is not the only possible theoretical stance. From a naturalistic standpoint, there is a strict correlation between cognitive capacities and consciousness, on the one hand, and neuronally-animated-bodies-in-the-world, on the other hand, whereas there is no evidence of freely wandering nervous systems displaying cognitive capacities, even in liquid media. One might naturally hypothesize, therefore, that cognitive capacities as well as consciousness have tightly coupled brain-
body systems as their core biological realizers, and not simply the brain alone.

Consider also the ontogeny of the individual organism. It is well known that the development of neural tissue depends on a complex pattern of interaction between proneural and nonneural tissues in the developing embryo. This interaction happens through selectively inhibiting and promoting the expression of a complex network of soluble and cell-associated molecules, such as growth factors, transcription factors, and membrane proteins (Glavic et al. 2004; Weinstein and Hemmati-Brivanlou 1999). Dorsal ectoderm, for instance, which is the origin of the entire central nervous system, differentiates into neural tissue in response to signaling from (nonneural) dorsal mesodermal tissue (the Spemann’s organizer in amphibians or the node in amniotes, such as the chick or the mouse) (De Robertis and Kuroda 2004; LaBonne and Bronner-Fraser 1999). Neural crest cells, which are the precursors of the peripheral nervous system, also give rise to bone tissue and smooth muscle, among other nonneural tissues (LaBonne and Bronner-Fraser 1999). Furthermore, peripheral factors, such as the sexual hormones testosterone and estrogen, as well as the hormone adipocyte-derived leptin, play a critical role in determining patterns of synaptogenesis and axon guidance, and therefore deeply influence the development of the nervous system (Lathe 2001; Morris, Jordan, and Breedlove 2004; Simerly 2005).

Thus, from a developmental perspective, it is not as if a commanding nervous system wraps itself with a body. Rather, it would be better to say that the body constructs a nervous system within itself. Clearly, the brain plays an undeniable role in enabling cognitive functions, as neuropsychological patients poignantly attest. Nevertheless, the brain is first and foremost responsible for the organism’s integrity while also being entirely dependent on that integrity. As we have seen, the brain plays this role by establishing and maintaining the internal regulatory processes and sensorimotor regularities that make up the homeodynamic domain that is the living body. As Piaget (1971) noted, this self-regulating domain shapes all cognitive processes and provides the ground state upon which any neural process, including those crucial for consciousness and subjectivity, can operate.

13.5 A Null Hypothesis for the Brain-in-a-Vat Thought Experiment: A Body in a World

The philosopher’s brain-in-a-vat thought experiment abstracts away from the ontogeny and evolution of adaptive brain-body-environment
interactions, and thus begins with a brain that already has a set of capacities or behavioral possibilities that transcend its actual structure. In other words, the thought experiment abstracts away from historical constraints on the biological realization of mind. Although some philosophers would argue that such historical constraints are relevant to whether a given biological structure at a given time instantiates or realizes a particular perceptual or cognitive state, almost all philosophers would argue that such constraints are irrelevant to the metaphysical question of whether a given biological structure at a given time instantiates or realizes subjectivity and consciousness.

We will not dispute this point. Rather, we wish to make a different observation. When we take into consideration the functional and structural interdependence of brain and body that evolutionary, developmental, physiological, and behavioral evidence suggest, then the philosopher’s naïve view of the brain in a vat simply will not do. The body is not just some kind of container, replaceable by a vat, that supports a commanding brain. The body is an active partner in the immensely complex and wide biological computations that the organism as a whole engages in while encountering an unpredictable world and maintaining its identity through time (Chiel and Beer 1997; Kutas and Federmeier 1998; Thompson and Varela 2001). Hence any “vat” capable of coupling with the brain in the requisite way must be able to duplicate these complex bodily processes.

We therefore propose the following null hypothesis for the brain-in-a-vat thought experiment: any vat capable of performing the necessary functions will have to be a surrogate body that both regulates and is regulated by the nervous system. In other words, the vat will have to exhibit a level of complexity at least as high as that of a living body with respect to bodily systems of life-regulation and sensorimotor coupling. Thus the entire system (vat plus brain) must satisfy these two basic requirements: (1) it must be energetically open and able to actively regulate the flow of matter and energy through it so as to control its own external boundary conditions (life-regulation), and (2) it must be capable of actively regulating its own sensorimotor interactions with the outside world (sensorimotor agency). In short, the entire system must amount to a biologically autonomous, sensorimotor agent. The null hypothesis is thus that a brain in a vat would in fact have to be a body in the world.

Given this null hypothesis, we can also advance the following more general hypothesis, the rejection of which entails the rejection of the enactive position:
The total realizer for consciousness (including subjectivity or phenomenal selfhood and specific states of phenomenal consciousness) is not the brain or some neural subsystem, but rather a whole living system, understood as an autonomous system made up of some crucial set of densely coupled neuronal and extraneuronal subsystems.

13.6 Putting Life Back into Consciousness

In conclusion, let us highlight two implications of our discussion that are relevant to the widely acknowledged explanatory gap between consciousness and the brain. First, given that consciousness is so clearly subordinate to the organism’s homeodynamic integrity, it may be more productive for research to proceed on the assumption that consciousness is a function of life-regulation processes involving dense couplings between neuronal and extraneuronal systems, rather than a function of neural systems alone (Thompson and Varela 2001). Second, mere neural correlates of consciousness will always leave an explanatory gap unless we know what role these neural correlates play in the context of the organism’s life-regulation and sensorimotor engagement with the world. The enactive approach aims to put life back into consciousness by building on these two points.

Notes

1. Searle thinks you really are a brain in a vat right now: “The vat is the skull and the ‘messages’ coming in are coming in by way of impacts on the nervous system” (1983, 230).

2. We are here setting aside the hard problem of what metaphysically constitutes consciousness.

3. Although the brain represents only approximately 2 percent of the total body mass, it is responsible for 20 percent of the energy from oxygen consumption in the body.

4. These include not only global arousal levels, but also the facilitation of selective patterns of regional synchronization within the general desynchronized cortical activity.

5. For the notion of biological autonomy, see Ruiz-Mirazo and Moreno 2004; Thompson 2007; and Varela 1979.
References


Embodiment or Envatment?


Embodiment or Envatment?


