The Conscious Id

Mark Solms (Cape Town)

Two aspects of the body are represented in the brain, and they are represented differently. The most important difference is that the brain regions for the two aspects of the body are associated with different aspects of consciousness. Very broadly speaking, the brainstem mechanisms derived from the autonomic body are associated with affective consciousness, and the cortical mechanisms derived from the sensorimotor body are associated with cognitive consciousness. Moreover, the upper brainstem is intrinsically conscious whereas the cortex is not; it derives its consciousness from the brainstem. These facts have substantial implications for psychoanalytic metapsychology because the upper brainstem (and associated limbic structures) performs the functions that Freud attributed to the id, while the cortex (and associated forebrain structures) performs the functions he attributed to the ego. This means that the id is the fount of consciousness and the ego is unconscious in itself. The basis for these conclusions, and some of their implications, are discussed here in a preliminary fashion.

Keywords: affect; cognition; conscious; ego; id; unconscious

1. Neuroanatomical representations of the body

At the 12th International Neuropsychoanalysis Congress (Berlin), on the topic of "Minding the Body," fresh light was cast on matters of fundamental interest to our field. Bud Craig, Antonio Damasio, Vittorio Gallese, Jaak Panksepp, and Manos Tsakiris, among others, summarized the current state of knowledge about *embodiment* in human neuropsychology (i.e., how the body is represented in the brain). In my concluding remarks to the congress, I pointed out that the speakers had referred to two different aspects of the body, without always distinguishing them. This can lead to confusion.

The first aspect of the body is neuroanatomically represented in somatotopic maps on the cortical surface, which are projections of sensory receptors on the surface of the body, relayed via modality-specific thalamic and cranial-nerve pathways. This aspect of body representation is conventionally equated with the cortical homunculus (the inverted little body-map that constitutes the primary somatosensory zone of the cortex).[1] But it does not coincide with somatosensory cortex alone; it includes the projection zones of all the sensory modalities, which consist in equivalent maps of the other sensory receptor organs (dark blue in Figure 1).

The "body image" arises not *in* but, rather, *from* these unimodal cortical maps. This first aspect of body representation should therefore be equated also with the processing networks that extend beyond the projection zones and converge in heteromodal association cortex (light blue in Figure 1).[2] We may call this aspect of body representation the *external body,* for short.

It is important to note that the corticothalamic mechanisms that represent the external body also represent other external objects—via the same perceptual modalities, in the same form. The external body is represented *as an object*. It is the form of the self that

[1] In fact, there are several such maps, each representing a different component of somatic sensation (touch, pain, vibration, temperature, etc.). The vestibulocerebellar system is also excluded from this simplified account.

[2] I do not mean to imply that the flow of information in this associative process is unidirectional. It is bidirectional, and, in fact, most of the connections lead in the other direction, from association to projection cortex. (See footnotes 4 and 12.)

Mark Solms: Professor of Neuropsychology, University of Cape Town, Cape Town, South Africa.
Correspondence to: Mark Solms, Department of Psychology, University of Cape Town, Rondebosch 7701, South Africa (email: Mark.Solms@uct.ac.za).
This article is based on: M. Solms, "Concluding Remarks," presented at the 12th International Neuropsychoanalysis Congress on 26 June 2011; M. Solms, "The Conscious Id," presented at the New York Psychoanalytic Institute on 5 November 2011; and M. Solms & J. Panksepp (2012), "The Id Knows More Than the Ego Admits."

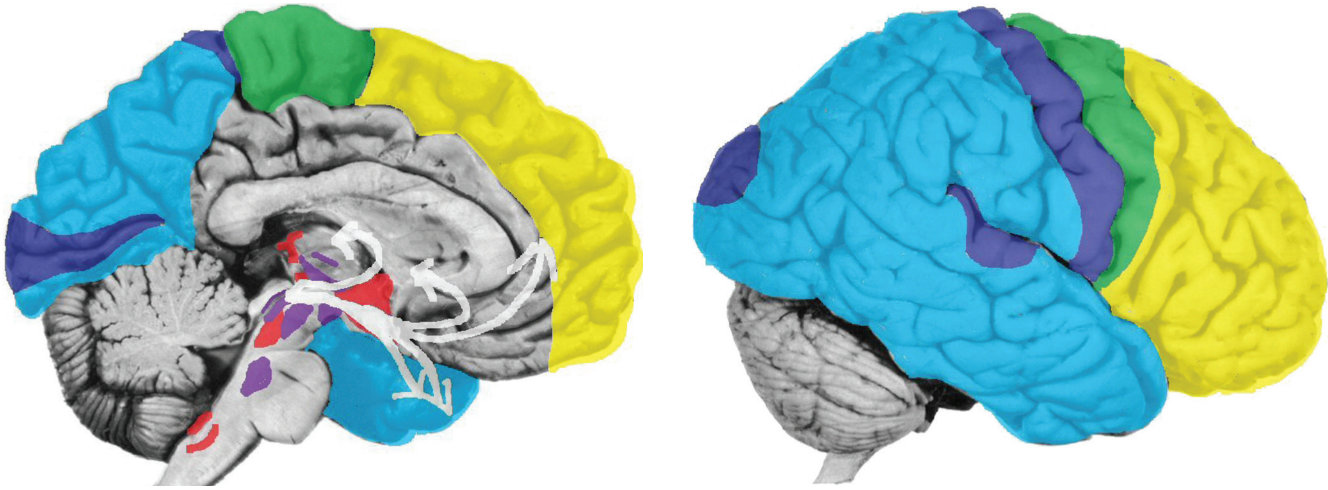

FIGURE 1. Lateral and medial views of the human brain. (Dark blue = sensory projection cortex; light blue = sensory association cortex; green = motor projection cortex; yellow = motor association cortex; red = autonomic nuclei; magenta = arousal nuclei; white = basic emotion circuits.)

one perceives when one looks outward, at a mirror for example. ("That thing is me; it is my body.") Other bodies are similarly represented.

It should be remembered that motor maps, too, contribute to the image of the external body.[3] The three-dimensional body image is generated not by heteromodal sensory convergence alone but also by movement. Movement produces sensation. The close relationship between movement and (kinesthetic) sensation is reflected in the anatomical proximity of the respective projection zones: the somatosensory and motor homunculi (green in Figure 1) form an integrated functional unit.

The second aspect of the body is its internal milieu, the autonomic body. This aspect of the body is barely represented on the cortical surface. It is represented much deeper and lower in the brain. The structures that represent this aspect of the body pivot around the hypothalamus, but they also include the circumventricular organs, parabrachial nucleus, area postrema, solitary nucleus, and the like (red in Figure 1; for review, see Damasio, 2010). Analogous to what I said above about the motor cortex in relation to exteroception, these interoceptive structures, too, not only monitor but also regulate the state of the body (homeostasis). We may call this aspect of body representation the *internal body*, for short.

Even at the level of the brainstem, the neural structures for internal body representation are surrounded by those for the external body, just as the sensorimotor body itself envelops the viscera.

[3] This applies both to one's own body and to other bodies (see discussion of "mirror neurons" below).

The internal body functions largely automatically, but it also arouses the external body to serve its vital needs in the external world. This is achieved through a network of upper-brainstem, diencephalic, and basal forebrain "arousal" structures (violet in Figure 1) known conventionally—but somewhat misleadingly—as the extended reticulo-thalamic activating system (ERTAS). This arousal system includes many long-axoned subsystems that release single neurotransmitters such as acetylcholine, noradrenaline, dopamine, serotonin, and histamine, as well as a variety of neuropeptides (for reviews, see Panksepp, 1998; Pfaff, 2006).

It is important to note that a hierarchical relationship exists between these two aspects of body representation. Although the flow of information (and therefore control) is both "bottom-up" and "top-down," the functional integrity of the cortex (external body) is contingent upon brainstem (internal body) activation. This hierarchical relationship involves consciousness. The arousal system associated with the internal body generates a different aspect of consciousness from that associated with external perception, and, moreover, *the internal aspect is prerequisite for the external aspect*. When endogenous consciousness is obliterated, exteroceptive awareness is obliterated too; however, the converse does not apply.[4]

The internal type of consciousness consists in *states*

[4] How exactly the exteroceptive varieties of conscious perception and cognition derive from ERTAS activation is uncertain, but some heuristic speculations are presented below. What is now widely accepted is the once radical notion that perceptual consciousness is endogenously generated; exteroceptive stimuli merely constrain and sculpt what is fundamentally a hallucinatory process (for reviews, see Blom & Sommer, 2012). Cf. footnote 2.

rather than *objects* of consciousness (cf. Mesulam, 2000). The internal body is not an object of perception unless it is externalized and presented to the classical senses; it is the *subject* of perception. It is the background state of *being* conscious. This is of paramount importance. We may picture this aspect of consciousness as the page upon which external perceptions are inscribed. The relationship between the two aspects of consciousness—the objects and the subject of perception—is also what binds the components of perception together; objects are always perceived by an experiencing subject (cf. the "binding problem").

It has recently been recognized that the state of the body-as-subject involves not only varying *levels* of consciousness (e.g., sleep/waking), but also varying *qualities* of consciousness (Damasio, 2010; Panksepp, 1998). The internal aspect of consciousness "feels like" something. Above all, the phenomenal states of the body-as-subject are experienced *affectively*. Affects do not emanate from the external sense modalities. They are states *of the subject*. These states are thought to represent the biological value of changing internal conditions (e.g., hunger, sexual arousal). When internal conditions favor survival and reproductive success, they feel "good"; when not, they feel "bad." This is evidently what conscious states are *for*. Conscious feelings tell the subject how well it is doing. At this level of the brain, therefore, consciousness is closely tied to homeostasis.

Affect may accordingly be described as an interoceptive sensory modality—but that is not all it is. Affect is an intrinsic property of the brain. This property is expressed in emotions, and emotions are, above all, peremptory forms of motor discharge. This reflects the fact that the changing internal conditions mentioned above are closely tied to changing external conditions. This is because, first, vital needs (represented as deviations from homeostatic set-points) can only be satisfied through interactions with the external world. Second, certain changes in external conditions have predictable implications for survival and reproductive success. Therefore, affects, although inherently subjective, are typically directed toward objects: "I feel like this *about that*" (cf. the philosophical concept of intentionality or "aboutness").

The keynote of affective consciousness is provided by the pleasure–unpleasure series, the motor expression of which is approach–withdrawal behavior. Feelings of pleasure–unpleasure—and the associated peremptory actions—are readily generated by stimulating a region of the ERTAS known as the periaqueductal grey (PAG). This ancient structure is found in all vertebrates. With increasing encephalization, however, a variety of more complex submodalities of affect and affective motivation appear, presumably through selective pressures arising from predictable conditions of major biological value. Thus, ascending from the PAG and into the limbic forebrain, which reciprocally provides descending controls, are various instinctual motivational circuits (white in Figure 1) that prepare mammalian organisms for situations of fixed biological value. These are known as the circuits for "basic emotions." They, too, are intrinsic to the brain and have inherent organization that is readily demonstrated by stimulating the relevant circuits (in all mammals, including humans).

There are several classifications of the basic emotions. The best-known taxonomy is that of Jaak Panksepp (1998), which recognizes (1) appetitive foraging, (2) consummatory reward, (3) freezing and flight, (4) angry attack, (5) nurturant care, (6) separation distress, and (7) rough-and-tumble play. The basic-emotion systems are given capitalized names—SEEKING, LUST, FEAR, RAGE, CARE, GRIEF, PLAY—to distinguish them from the equivalent colloquial usages. It is important to note that each of these circuits generates not only stereotyped actions, but also specific feelings and motivations, such as curiosity, sensuality, trepidation, anger, affection, sorrow, and joy. The brain circuits for the basic emotions are conserved across the mammalian series, and they admit of considerable chemical specificity.

To be clear: the basic emotions enumerated above do not exhaust the range of human affectivity. What distinguishes them is their *instinctual* nature. There are whole classes of simpler affects, such as *homeostatic affects*, which give expression to vegetative drives (e.g., hunger and thirst), and *sensory affects*, which respond automatically to certain stimuli (e.g., surprise and disgust), not to mention the infinite variety of hybrid forms generated when any of these affects blends with cognition (see Panksepp, 1998).

2. Metapsychological representations of the body

Having reviewed the two ways in which the body is represented in the brain, it is easy to recognize the neurological equivalents of the two major mental systems that Sigmund Freud distinguished in his metapsychology. The external body corresponds to the "ego," the internal body to the "id" (see Figure 2; cf. Figure 1).

Freud himself said as much. About the bodily derivation of the "ego," Freud wrote this:

> The ego is first and foremost a bodily ego; it is not merely a surface entity, but is itself the projection of a

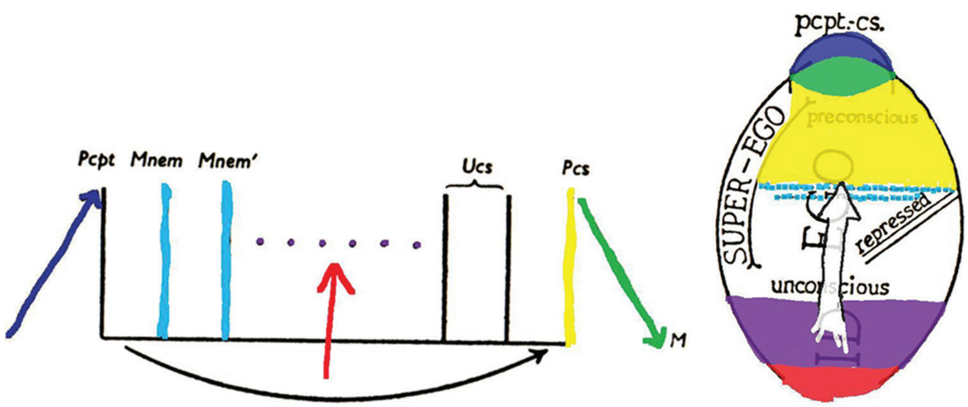

FIGURE 2. Freud's classical models of the mind, color-coded to illustrate the metapsychological correlates of the anatomical regions identified in Figure 1

> surface. If we wish to find an anatomical analogy for it we can best identify it with the "cortical homunculus" of the anatomists, which stands on its head in the cortex, sticks up its heels, faces backwards and, as we know, has its speech-area on the left-hand side. [Freud 1923, p. 26]

He elaborated:

> The ego is ultimately derived from bodily sensations, chiefly from those springing from the surface of the body. It may thus be regarded as a mental projection of the surface of the body, besides, as we have seen above, representing the superficies of the mental apparatus.

The whole fabric of the ego is derived from this bodily ego—that is, from memory traces of external perception (Figure 2), the associative activation of which gives rise to all cognition (see Sections 6–9).

About the bodily derivation of the "id," Freud wrote:

> The id, cut off from the external world, has a world of perception of its own. It detects with extraordinary acuteness certain changes in its interior, especially oscillations in the tension of its instinctual needs, and these changes become conscious as feelings in the pleasure–unpleasure series. It is hard to say, to be sure, by what means and with the help of what sensory terminal organs these perceptions come about. But it is an established fact that self-perceptions—coenaesthetic feelings and feelings of pleasure–unpleasure—govern the passage of events in the id with despotic force. The id obeys the inexorable pleasure principle. [Freud 1940, p. 198]

The word "instinctual" here is a mistranslation of the German word *Triebe*. A *Trieb* is a "drive." Freud clearly defined what he meant by the term:

> An "instinct" [*Trieb*] appears to us as a concept on the frontier between the mental and the somatic, as the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demand made upon the mind for work in consequence of its connection with the body. [Freud 1915a, pp. 121–122]

It is thus evident that Freud himself readily localized the bodily derivations of the ego and the id. His conception of the mental apparatus was always embodied—that is, it was tethered to the body at its perceptual/motor and "instinctual" ends (Figure 2). I have only added anatomical detail here. I have also clarified that instincts consist in more than interoceptive perception; they are intrinsic emotional stereotypes. But Freud did recognize the instinctual nature of what are now called basic emotions:

> And what is an affect in the dynamic sense? It is in any case something highly composite. An affect includes in the first place particular motor innervations or discharges and secondly certain feelings; the latter are of two kinds—perceptions of the motor actions that have occurred and the direct feelings of pleasure and unpleasure which, as we say, give the affect its keynote. But I do not think that with this enumeration we have arrived at the essence of an affect. We seem to see deeper in the case of some affects and to recognize that the core which holds the combination we have described together is the repetition of some particular significant experience. This experience could only be a very early impression of a very general nature, placed in the prehistory not of the individual but of the species. [Freud 1916–17, p. 395]

Notwithstanding Freud's tendency to describe phylogenetic associations as if they were literally remembered, he recognized—as later would Panksepp (1998)—that the basic emotions are *innate* mental organizations. (This contrasts with the James–Lange theory: James, 1890; Lange, 1885.)

In short: it is easy to recognize a functional equivalence between the brain mechanisms for the external body and Freud's bodily ego, on the one hand, and between those for the internal body and Freud's id instincts, on the other. This applies equally to the interdependent hierarchical relationship that pertains between them: there can be no cortical consciousness without brainstem consciousness; there can be no ego without id.

3. The corticocentric fallacy

This close parallelism reveals a stark contradiction between the current concepts of affective neuroscience and those of Freud.

To fully expose the contradiction, I need to point out that Freud never questioned a classical assumption of nineteenth-century neuroanatomists—namely, that consciousness was a *cortical* function:

> What consciousness yields consists essentially of perceptions of excitations coming from the external world and of feelings of pleasure and unpleasure which can only arise from within the mental apparatus; it is therefore possible to assign to the system Pcpt.-Cs. a position in space. It must lie on the borderline between inside and outside; it must be turned towards the external world and must envelop the other psychical systems. It will be seen that there is nothing daringly new in these assumptions; *we have merely adopted the views on localization held by cerebral anatomy, which locates the "seat" of consciousness in the cerebral cortex*—the outermost, enveloping layer of the central organ. Cerebral anatomy has no need to consider why, speaking anatomically, consciousness should be lodged on the surface of the brain instead of being safely housed somewhere in its inmost interior. [Freud, 1920, p. 24; emphasis added]

To be sure, Freud recognized that consciousness also entailed "feelings of pleasure and unpleasure which can only arise from within the mental apparatus" (1920). He even suggested that this aspect defined the biological purpose of consciousness (Freud, 1911, p. 220). That is why Antonio Damasio was moved to say that "Freud's insights on the nature of affect are consonant with the most advanced contemporary neuroscience views" (Damasio, 1999a, p. 38). But it is clear from the preceding quotation that even the internal aspect of consciousness was, for Freud, "lodged on the surface of the brain." Here he states this view even more explicitly:

> The process of something becoming conscious is above all linked with the perceptions which our sense organs receive from the external world. From the topographical point of view, therefore, it is a phenomenon which takes place in the outermost cortex of the ego. It is true that we also receive information from the inside of the body—the feelings, which actually exercise a more peremptory influence on our mental life than external perceptions; moreover, in certain circumstances the sense organs themselves transmit feelings, sensations of pain, in addition to the perceptions specific to them. Since, however, these sensations (as we call them in contrast to conscious perceptions) also emanate from the terminal organs *and since we regard all these as prolongations or offshoots of the cortical layer*, we are still able to maintain the assertion made above. The only distinction would be that, as regards the terminal organs of sensation and feeling, the body itself would take the place of the external world. [Freud, 1940, pp. 161–162; emphasis added][5]

[5] Freud's localization of consciousness underwent many vicissitudes. Initially he made no distinction between perceptual and affective consciousness (Freud, 1894). Rather, he distinguished between *memory traces of perception* ("ideas") and the *energy that activates them*. This distinction coincided with the conventional assumptions of British empiricist philosophy, but Freud interestingly described the activating energy as "quotas of affect," which are "spread over the memory-traces of ideas somewhat as an electric charge is spread over the surface of a body" (Freud, 1894, p. 60). Strachey (1962, p. 63) described this as the "most fundamental of all [Freud's] hypotheses." There is every reason to believe that Freud envisaged such activated memory traces of "ideas" as *cortical* processes. In his more elaborated (1895) "Project" model, he explicitly attributed consciousness to a subsystem of cortical neurons (the ω system), which he located at the *motor* end of the forebrain. This location enabled consciousness to register discharge (or lack thereof) of the energy that accumulated over the memory traces (the ψ system) from both endogenous and sensory sources. (Note that from 1895 onward Freud described mental energy as being unconscious in itself; it was no longer described as a "quota of affect.") Consciousness, which Freud now divided into two forms, arose from *the manner in which mental energy excited the* ω *neurons*. It gave rise to *affective* consciousness when differences in the quantitative level of energy in the ψ system (caused by degrees of motor discharge) were registered in ω as pleasure–unpleasure; it gave rise also to *perceptual* consciousness when differences in qualitative aspects of exogenous energies (e.g., wavelength or frequency) derived from the different sense organs were transmitted, via perceptual (φ) neurons, through the memory traces of ideas (ψ), onto ω. In an 1896 revision of this "Project" model, Freud moved the ω neurons to a position between φ and ψ, and he simultaneously acknowledged that all energy in the mental apparatus was endogenously generated; energy did not literally enter the apparatus through the perceptual system. (Freud seemed to forget this later—e.g., 1920.) In *The Interpretation of Dreams* (1900), however, Freud reverted to the "Project" arrangement and again located the perceptual and consciousness systems at opposite ends of the mental apparatus. His indecision in this respect seems to have derived mainly from the fact that the perceptual (sensory) and consciousness (motor) systems formed an integrated functional unit, since motor discharge necessarily produces perceptual information (cf. the contiguous location of the somatosensory and motor homunculi; Figure 1). Freud accordingly settled (in 1917) on a hybrid localization of the perceptual and consciousness systems. In this final arrangement, φ (renamed "Pcpt." in 1900) and ω ("Cs.") were combined into a single functional unit, the system "Pcpt.–Cs." (Figure 2). At this point, Freud clarified that the Pcpt.-Cs. system is really a single system, which is *excitable from two directions*: exogenous stimuli generate perceptual consciousness, endogenous stimuli generate affective consciousness. Freud also retreated from the notion that affective consciousness registers

In making this assumption, Freud was following a long tradition, which continues to this day, even among some eminent cognitive and behavioral neuroscientists. Consider for example the following remark made by Joseph LeDoux:

> When electrical stimuli applied to the amygdala of humans elicit feelings of fear (see Gloor, 1992), it is not because the amygdala "feels" fear, but instead because the various networks that the amygdala activates ultimately provide *working memory* with inputs that are *labeled* as fear. This is all compatible with the Freudian notion that conscious emotion is the awareness of something that is basically unconscious. [LeDoux, 1999, p. 46; emphasis added]

Such "corticocentric" theorists simply assume that all consciousness is cortical, which implies that affective states generated deeper in the brain can only become conscious when they are read out (or "labeled") in the higher reaches of working memory. As we shall see below, this view is sharply contradicted by all available evidence. The latest influential representative of the corticocentric tradition is Bud Craig (2009). Craig even believes there is a cortical projection zone for the internal body, in the posterior insula. He equates this cortical region with the body-as-subject, the primary sentient "self"—precisely the function that I have attributed, on the basis of a different research tradition, to the upper brainstem and limbic system.

4. Consciousness without cortex

Recent research demonstrates unequivocally that the corticocentric view of consciousness (as the seat of the sentient self) is mistaken. Consider the following interview, reported at our Berlin congress by Damasio (and since published in Damasio, Damasio, & Tranel, 2012), concerning a patient in whom the insula was *totally obliterated bilaterally* by herpes simplex encephalitis. According to Craig's view, this patient should lack phenomenal selfhood; he should lack the very page upon which experience is inscribed. But this is not the case:

Q: "Do you have a sense of self?"

A: "Yes, I do."

Q: "What if I told you that you weren't here right now?"

A: "I'd say you've gone blind and deaf."

Q: "Do you think that other people can control your thoughts?"

A: "No."

Q: "And why do you think that's not possible?"

A: "You control your own mind, hopefully."

Q: "What if I were to tell you that your mind was the mind of somebody else?"

A: "When was the transplant, I mean, the brain transplant?"

Q: "What if I were to tell you that I know better than you know yourself?"

A: "I would think you're wrong."

Q: "What if I were to tell you that you are aware that I'm aware?"

A: "I would say you're right."

Q: "You are aware that I am aware?"

A: "I am aware that you are aware that I am aware."

This case does not disprove the entire corticocentric theory of consciousness; it disproves only Craig's (insular) version of the theory. But what about the rest of the cortex?

In animal models, the removal of cortex was long ago shown to have no effect on behavioral proxies of consciousness, such as sleep/waking and instinctual-emotional actions. Indeed, not only are the rewarding and punishing effects of subcortical brain stimulation demonstrably preserved in decorticated animals, they are actually enhanced, presumably due to the release of "top-down" cortical inhibition of emotional consciousness (Huston & Borbely, 1974).

The most striking evidence to emerge in recent years from human research relevant to this broader question concerns a condition called hydranencephaly, in which the cerebral cortex is destroyed *in utero* (usually due to infarction of the entire anterior cerebral circulation). The cortex is absorbed and replaced with cerebrospinal fluid (Figure 3). Autopsy studies reveal that although fragments of cortex may be preserved in such cases, they are disconnected from the thalamus due to destruction of the linking white matter. The surviving cortical fragments are also gliotic and are therefore completely nonfunctional. This is confirmed by the clinical observation that, although some visual

---

the quantitative "level" of excitation within the ψ system, and he suggested instead that it—like perceptual consciousness—registers something qualitative, like wavelength (i.e., fluctuations in the level of energy within the Pcs system over a unit of time; see Freud, 1920). The main thing to notice in this brief history of Freud's localization of consciousness is that it was from first to last conceptualized as a cortical process (although Freud did seem to have fleeting doubts about this at times; e.g., 1923, p. 21). (See Solms, 1997, for a first intimation that something was wrong with Freud's superficial localization of the internal (affective) surface of the system Pcpt.-Cs.)

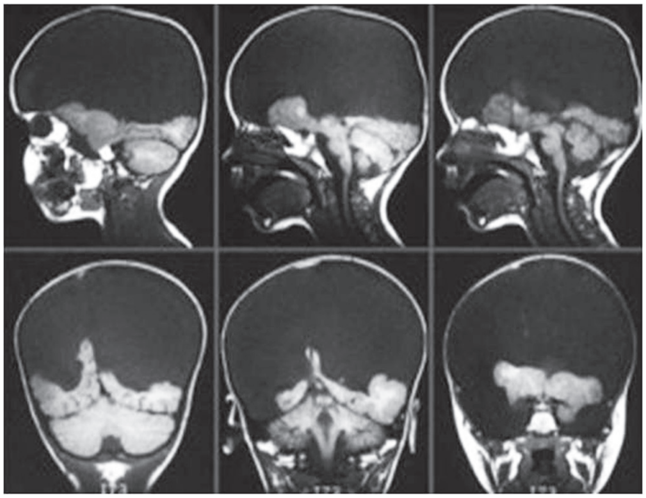

FIGURE 3. A typical hydranencephalic brain. (Reproduced with permission of the American College of Radiology. No other representation of this material is authorized without expressed, written permission from the American College of Radiology.)

cortex may be preserved, the patients are blind (Merker, 2007).

They are blind (etc.),[6] but they are not unconscious. These children display normal sleep/waking cycles. They also suffer absence seizures, in which the parents have no trouble recognizing the lapses of consciousness and when the child is "back" again. This is weighty evidence in favor of the view that they are conscious. The detailed clinical reports of Shewmon, Holmse, and Byrne (1999) provide further evidence that these children not only qualify as conscious by the standard behavioral criteria of the Glasgow Coma Scale, but they also show vivid emotional reactions (see Figure 4, for example, which illustrates the reaction of a hydranencephalic girl when her baby brother is placed in her arms):

> They express pleasure by smiling and laughter, and aversion by "fussing," arching of the back and crying (in many gradations), their faces being animated by these emotional states. A familiar adult can employ this responsiveness to build up play sequences predictably progressing from smiling, through giggling, to laughter and great excitement on the part of the child. [Merker, 2007, p. 79]

They also show associative emotional learning:

> [They] take behavioral initiatives within the severe limitations of their motor disabilities, in the form of instrumental behaviors such as making noise by kicking trinkets hanging in a special frame constructed for the purpose ("little room"), or activating favorite toys by switches, presumably based upon associative learning of the connection between actions and their effects. Such behaviors are accompanied by situationally appropriate signs of pleasure and excitement on the part of the child. [p. 79]

FIGURE 4. Expression of pleasurable emotion in a young hydranencephalic girl. (Reproduced with permission of the child's mother, with thanks to Bjorn Merker.)

In short: although there is in these children significant degradation of the types of consciousness that are normally associated with representational perception and the cognition derived from it, there can be no doubt that they are conscious, both quantitatively and qualitatively. They are not only awake and alert, they also experience and express a full range of instinctual emotions. The primary (affective) self is, in short, *present*. The fact that cortex is absent in these cases proves that affective consciousness is both generated and felt subcortically. This contradicts the theoretical assumptions of LeDoux and Craig quoted above, and those of Freud.

Sadly, in this respect Freud seems to have paved the way for the conflation of consciousness with cortical monitoring, thereby prematurely relegating unmonitored instinctual processes to the "unconscious" category. It is now clear that instinctual processes are conscious in themselves.

5. All consciousness is endogenous

The state of consciousness as a whole is generated in the upper brainstem. We have known this for many years. A mere decade after the death of Freud, Moruzzi and Magoun (1949) first demonstrated that consciousness, as measured by EEG activation, is generated in a part of the upper brainstem then called the "reticular activating system." Total destruction of exteroceptive structures had no impact on the intrinsic

[6] They lack perceptual *consciousness*. This does not mean they cannot process exteroceptive *information* via subcortical pathways. Consciousness is not prerequisite for perception (cf. "blindsight"). This point is important for my argument (below) to the effect that the ego is unconscious in itself.

consciousness-generating properties of the brainstem system (e.g., sleep/waking). Moruzzi and Magoun's conclusions (in cats) were quickly confirmed (in humans) by Penfield and Jasper (1954), who recognized in absence seizures (mentioned above) "a unique opportunity to study the neuronal substratum of consciousness" (p. 480). Their extensive studies led them to the conclusion that paroxysmal obliterations of consciousness could only be reliably triggered at an upper-brainstem site (which they termed the "centrencephalic system"). They were also impressed by the fact that removal of large parts of human cortex under local anesthetic, even total hemispherectomy, had limited effects on consciousness. Cortical removal did not interrupt the presence of the sentient self, of being conscious; it merely deprived the patient of "certain forms of information" (Merker, 2007, p. 65). Lesions in the upper brainstem, by contrast, totally and rapidly destroyed consciousness, just as the induced seizures did. These observations demonstrate a point of fundamental importance: *all* consciousness ultimately derives from upper-brainstem sources. In sharp contradiction to the corticocentric assumption, cortical varieties of consciousness actually depend on the integrity of subcortical structures, not the other way round.

The classical observations that underpinned this important conclusion have stood the test of time, with greater anatomical precision being added (for review, see Merker, 2007). Significantly, the PAG appears to be a nodal point in the "centrencephalic system." This is the smallest region of brain tissue in which damage leads to total obliteration of consciousness. That observation underscores the single fact that *has* changed in modern conceptions of this system: the deep structures that generate conscious state are not only responsible for the level but also for the core quality of subjective being. Conscious states are inherently *affective.* It is this realization that is now revolutionizing consciousness studies (Damasio, 2010; Panksepp, 1998).

The classical conception is turned on its head. Consciousness is not generated in the cortex; it is generated in the brainstem. Moreover, consciousness is not inherently perceptual; it is inherently affective. And in its primary manifestations, it has less to do with cognition than with instinct. In terms of the parallels drawn in Section 2, the conclusion is inescapable: *consciousness is generated in the id*, and the ego is fundamentally unconscious. This has massive implications for our conceptualization of the ego and all that flows from it, such as our theories of psychopathology and clinical technique. It was, after all, the essence of the "talking cure" that words, being ego memory-traces derived from external perception and therefore capable of consciousness, must be attached to the deeper processes of the mind (which are unconscious in themselves) before they can be known by the subject.

6. Mental solids

What, then, does cortex contribute to consciousness? The answer to this question will shed new light on the metapsychological status of the ego. It is clear from the facts just reviewed that consciousness attached to exteroceptive information processing is not intrinsic to the cortex but, rather, derives from brainstem sources. Cortex without a brainstem can never be conscious. Perceptual processing therefore does not require consciousness, as is amply demonstrated by the vast abilities of the "cognitive unconscious" (for review, see Kihlstrom, 1996).

Moreover, much of what we have traditionally thought to be "hard-wired" in cortical processing is actually *learnt*. This has been well demonstrated by the research of Mriganka Sur, which shows, for example, that redirecting visual input from occipital cortex to auditory cortex (in ferrets) leads to reorganization of the latter tissue to support completely competent vision (for review, see Sur & Rubinstein, 2005). Cortical perception, therefore, no less than cortical cognition, is rooted in *memory* processes. Indeed, as far as we know, all cortical functional specializations are acquired. The columns of cortex are initially almost identical in neural architecture, and the famous differences in Brodmann's areas probably arise from use-dependent plasticity (following the innate patterns of subcortical connectivity). Cortical columns resemble the random-access memory (RAM) chips of digital computers.

The answer to our question, "What does cortex contribute to consciousness?", then, is this: it contributes representational memory space. This enables cortex to *stabilize* the objects of perception, which in turn creates potential for detailed and synchronized processing of perceptual images. This contribution derives from the unrivalled capacity of cortex for *representational* forms of memory (in all of its varieties, both short- and long-term).[7] Based on this capacity, cortex transforms the fleeting, wavelike states of brainstem activation into "mental solids." It generates *objects*. Freud called them "object-presentations" (which, ironically, predominate in what he called the "system unconscious").

Such stable representations, once established through learning, can be activated both externally and

[7] It will be noted that this representational capacity derives from the topological "mapping" of the external body, described in Section 1.

internally, thereby generating objects not only for perception, but also for cognition (perception involves recognition).[8] To be clear: the cortical representations are unconscious in themselves; however, when consciousness is *extended onto* them (by "attention"),[9] they are transformed into something both conscious *and* stable, something that can be *thought* in working memory. (It is no accident that we describe the consciousness of everyday experience as working *memory*.) The activation by brainstem consciousness-generating mechanisms of cortical representations thus transforms consciousness from affects into objects.[10] The transformation is, however, never complete: conscious representations must still be experienced *by* a subject, and working memory typically contains elements of both cognitive and affective consciousness. Remarkably, most cognitive theorists just ignore the affect.

But why does "working memory" have to be conscious? I have already explained why any form of perceptual representation is infused with consciousness: it gives *valence* to the representations ("I feel like this about that"). Although this formulation derives from Damasio's conception of *The Feeling of What Happens* (1999b), it also invokes Freud's (1895) earlier notion that the forebrain is a "sympathetic ganglion"—that is, that perceptual learning only exists because it serves vital (survival and reproductive) needs. Learning entails the establishment of associations between interoceptive drives and exteroceptive representations, guided by the feelings that are generated in such encounters.[11] This enables the subject to feel its way through novel situations. The affective "presence" of the subject is required to do so.

If such encounters are to issue in more than stereotyped instinctual responses, they also require *thinking*. And thinking necessarily entails *delay*. This (delay) function is rooted first and foremost in the stability of cortical representations, which enables them to be "held in mind." The prototype for this in Freud's metapsychology was "wishful cathexis," which entails a representation of the wished-for object being used to guide ongoing behavior. In the first instance, however, such volitional behavior is regulated directly by instinct (by Freud's "pleasure principle" and its accompanying "primary-process" mode of cognition). Instinctual motivations are initially objectless (cf. the SEEKING concept of Panksepp; Wright & Panksepp, 2012), but sympathetic learning rapidly leads to *remembered* objects of desire coming to mind (cf. the "wanting" concept of Berridge, 1996). In other words, biologically valenced (wished-for, feared, etc.) objects of past experience are rendered conscious by dint of their "incentive salience" (which is ultimately determined by their biological meaning in the pleasure–unpleasure series—the very basis of consciousness). In this way, if left to its own devices, the pleasure principle would produce what Freud termed hallucinatory wish-fulfillments (the prototype of primary-process cognition).[12] It is important to note that conscious thinking, in itself, therefore does not necessarily entail what Freud called "secondary-process" cognition. Hallucinatory wish-fulfillment—Freud's prototype of "primary-process" thinking—is a *conscious* form of thinking, albeit a very primitive form.

Hence the evolutionary and developmental pressure to constrain incentive salience in perception through *prediction-error coding* (this is Freud's "reality principle"), which places constraints on motor discharge. Such error-coding must be regulated at bottom by the homeostatic function of affective consciousness, which determines the biological value of all objects of attention (cf. Freud's "constancy principle"). The resultant inhibition—which perforce occurs at the *motor* (frontal) end of the apparatus, where outputs *must be sequenced over time*—requires tolerance of frustrated emotions. This frustration, which gives rise to fresh thinking, and thus to new learning, secures more efficient biological satisfaction in the long run. (This is Freud's concept of "binding.")

Sequencing over time, which requires thinking *ahead* (i.e., virtual action, or action programming) defines the essence of the executive function of "working memory," in the sense that we generally theorize it today. Freud would have called this executive function "secondary-process" thinking (which he also

[8] Cf. Edelman's memorable phrase, "the remembered present."

[9] Cf. Freud's description of the process: "Cathectic innervations are sent out and withdrawn in rapid periodic impulses from within into the completely pervious system Pcpt.-Cs. So long as that system is cathected in this manner it receives perceptions (which are accompanied by consciousness) and passes the excitation onwards to the unconscious mnemic systems; but as soon as the cathexis is withdrawn, consciousness is extinguished and the functioning of the system comes to a standstill. It is as though the unconscious stretches out feelers, through the medium of the system Pcpt.-Cs., towards the external world and hastily withdraws them as soon as they have sampled the excitations coming from it" (Freud, 1925a, p. 231). Note that Freud's "feelers" of perception are *unconscious* until they reach the cortical system Pcpt.-Cs.

[10] It is very important to note that it also transforms the unconscious representations themselves, through the process of "reconsolidation." Indeed, representations only become conscious to the extent that corticothalamic predictive models of them are uncertain (i.e., subject to revision; see below).

[11] Higher-order associations (between the representations) are considered below.

[12] See footnote 2. Cf. Friston (2012): "Neuronal connections encode (model) causal connections that conspire to produce sensory information."

conceptualized as "experimental action"). But secondary-process thinking also entails other aspects of cortical functioning that we have not yet fully considered (see Section 8).

This, then, is the essential function of cortex. It generates stable, representational "mental solids" that, when activated (or "cathected") by affective consciousness, enable the id to *picture* itself in the world and to think. But mental solids also threaten to obscure all else from view, even in primary-process cognition. One is reminded of Plato's cave.

7. A surprise

Freud's secondary process rests upon "binding" of "free" drive energies.[13] Binding (i.e., inhibition) creates a reservoir of tonic activation that can be utilized for greatly enhancing the functions of thinking, just described, which Freud attributed to the ego. In fact, Freud's earliest conception of the ego defined it as a network of "constantly cathected" neurons that exerted collateral inhibitory effects on each other (Freud, 1895). This prompted Carhart-Harris and Friston (2010) to equate Freud's ego reservoir with the "default mode network" of contemporary cognitive neuroscience. Be that as it may, Karl Friston's work is grounded in the same Helmholtzian energy concepts as Freud's (see Friston, 2010). His model (in terms of which prediction-error or "surprise"—equated with free energy—is minimized through the consequent encoding of better models of the world, resulting in better predictions) is entirely consistent with Freud's. His model beautifully reconceptualizes Freud's "reality principle" in computational terms, with all the advantages this entails for quantification and experimental modeling. On this view, *free energy is untransformed affect*—energy released from the bound state, or blocked from entering the bound state, due to prediction errors (violations of the reality principle).

It is of the outmost importance to note that in Friston's model prediction error (mediated by surprise), which increases incentive salience (and therefore consciousness) in perception and cognition, *is a bad thing* biologically speaking. The more veridical the brain's predictive model of the world, then the less surprise, the less salience, the less consciousness, the more *automaticity*, the better. One is reminded of Freud's "Nirvana principle," which he took to be the ultimate goal of mental life.

The very purpose of the reality principle, which first gave rise to secondary-process (inhibited) cognition, is automaticity, *which obviates the need for consciousness* (it obviates the need for the subject to "feel its way" through situations).[14] This in turn suggests that the ideal of cognition is to forego representational (and therefore cortical) processing and replace it with associative processing—to shift from episodic to procedural modes of functioning (and therefore, presumably, from cortex to dorsal basal ganglia). It appears that consciousness in cognition is a temporary measure: a compromise. But with reality being what it is—always uncertain and unpredictable, always full of surprises—there is little risk that we shall in our lifetimes actually reach the zombie-like state of Nirvana that we now learn, to our surprise, is what the ego aspires to.

8. Words and things

Before we can leave the subject of cortex, I must point out that secondary-process thinking entails important features that were left implicit in the previous sections, especially regarding delayed response. These features are attributable to something other than representational and inhibitory capacity alone.

The wished-for object presentations that literally "come to mind" in primary-process (hallucinatory) thinking are, according to Freud, re-represented at a higher level in secondary-process thinking. He called this level of representation "word-presentation." Freud thought that the value of words was that they, like all cognitive presentations, are derived from perception (in this case, mainly hearing) and are therefore capable of consciousness. This is the nub of the role that words play in the "talking cure." But because words have the additional capacity of representing *relations between* the concrete objects of thought ("which is what specially characterizes thoughts, and cannot be given visual expression"; Freud, 1923, p. 21), they also render *abstract* cognition "declarative."

The principal value of words, therefore, is not that they enable us to *render conscious* the inchoate processes of the id (which Freud thought was unconscious); what is most important about words is their capacity to represent the relations between things, to *re*-represent them abstractly. This enables us to think

[13] Freud's psychological distinction between bound and free energy almost certainly derived from the physical distinction between potential and kinetic energy. This resolves an aspect of the "mind–body problem" (the supposed violation of Helmholtz's energy-conservation law). By definition, thinking (bound energy) has no effects until it is discharged in action.

[14] One is tempted to reverse Freud's famous dictum and say that "a memory-trace arises instead of consciousness" (cf. Freud, 1920, p. 25).

about things, as opposed to simply think things (to think in images). This underpins the all-important "third-person" perspective, to which we shall return shortly.

Something else important about words is syntax. The nineteenth-century psychology of words (Freud, 1891) long ago evolved into a psychology of *language.* The structure of language facilitates cortical programming of the delayed and sequenced responses discussed above: "first I will do this, then I will do that." The capacity of language to hold future-oriented programs in mind defines the modus operandi of the executive function of working memory (cf. "inner speech"). This is a special case of the capacity of words to represent the relations between things and, thereby, to render abstractions conscious. In short, words enable us to think about the relations between things both in space *and in time*. This greatly enhances the delayed-response mechanism and surely defines the essence of what Freud called "secondary-process" thinking. It is therefore important to remember that in Freud's second topographic model (Freud, 1923), he himself recognized that the capacity of the ego for secondary-process cognition was its defining feature—not its capacity for representational consciousness.

9. The reflexive ego, the superego

I said in Section 1 that the external body is made of the same perceptual stuff as other objects, that the bodily ego is inscribed in much the same way as other objects on the page of consciousness. It is a stabilized *representation* of the subject of consciousness—an object, a mental solid—experienced *by* the subject of consciousness. This primary subject of consciousness (the body-as-subject) is the id. It is important to recognize that the bodily "self" is an *idea*, albeit an everyday one.[15] It is a learnt representation of the self.

To this object-presentation we have to add a further complication called "Mark Solms"—the word-presentation—which is neither really me nor an animated picture of me, but, rather, an abstraction. In order to do so, I must say a bit more about the relationship between subjective "presence" of the id and objective representation of the body.

The subject of consciousness identifies itself with its external body (object-presentation) in much the same way as a child projects itself into the animated figure it controls in a television game. The representation is rapidly invested with a sense of self, although it is not really the self.

Here is a striking experiment that vividly illustrates the counterintuitive relation that actually exists between the subjective self and its external body. Petkova and Ehrsson (2008) report a series of "body swap" experiments in which cameras mounted over the eyes of other people, or mannequins, transmitting images from that viewpoint to video-monitoring goggles mounted over the eyes of the experimental subjects, rapidly created the illusion in the experimental subjects that the other person's body or the mannequin was their own body. This illusion was so compelling that it persisted even when the projected subjects shook hands with their own bodies. The existence of the illusion was also demonstrated objectively by the fact that when the other (illusory own) body and the (real) own body were both threatened with a knife, the fear response—the "gut reaction" of the internal body (measured by heart rate and galvanic skin response)—was greater for the illusory body.

The well-known "rubber hand illusion" (Botvinick & Cohen, 1998), which Tsakiris (2011) described in Berlin, demonstrates the same relation between the self and the external body, albeit less dramatically. The anatomical basis of such phenomena (which place Freud's theory of "narcissism" on a promising new empirical footing) may be linked with well-known fMRI and other findings to the effect that the topographic arrangement of somatosensory and motor cortical homunculi (the acknowledged locus of Freud's "bodily ego") can be readily manipulated and extended, even to include inanimate tools (for review, see Maravita & Iriki, 2004). We are reminded that cortex is nothing but random-access memory.

The learnt nature of the external body is further demonstrated by some striking "mirror-neuron" phenomena. Gallese (2011) reminded us at the Berlin congress that mirror neurons fire in the same way regardless of whether a movement is performed by the self or by the other (see also Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). How then does the self tell the difference—how does it know whether such movements are being performed by "me" or not? Evidently something has to be *added* to the motor cortical (mirror-neuron) activity for this distinction to be made. It appears that this "something" is concurrent frontal inhibition (which suppresses posterior insula activation). Gallese reported that schizophrenic patients cannot adequately differentiate between their own movements and those of others, for the reason that this concurrent inhibition is lacking in them (Ebisch et al., 2012).

[15] It is an everyday idea *in health*, which can disintegrate in pathological states (e.g., out-of-body experiences, autoscopic phenomena, ideas of reference).

This again demonstrates, first, that the external body is not a subject but an object, and second, that it is perceived in the same register as other objects.

In making this distinction between "me" and "not-me," the role of *words* in reflexive consciousness (a.k.a. secondary consciousness, access consciousness, declarative consciousness, autonoetic consciousness, higher-order thought, etc.), described above, is pivotal. This abstract level of re-representation enables the subject of consciousness to transcend its concrete "presence" and thus *to separate itself as an object from other objects*.[16] The process seems to unfold over three levels of experience: (1) the affective or phenomenal level of the self as subject, a.k.a. first-person perspective; (2) the perceptual or representational level of the self as object, a.k.a. second-person perspective; (3) the abstracted or re-representational level of the self as object in relation to other objects, a.k.a. third-person perspective.

The self of everyday experience tends ordinarily to think *about* itself from the third-person perspective, in relation to other objects, in such banal situations as "I willed that movement" (not the other person). We can only conclude that the self of everyday experience is largely an abstraction. This reveals the power of words.

The unrecognized gap between the primary subjective self and the re-representational "declarative" self causes much confusion. Witness the famous example of Benjamin Libet recording a delay of up to 400 ms between the physiological appearance of premotor activation and the voluntary decision to move. This is typically interpreted to show that free will is an illusion, when in fact it shows only that the verbally mediated, reflexive re-representation of the declarative self initiating a movement occurs somewhat later than the affective (primary) self actually initiating it. Such confusion would be avoided if we recognized that the self unfolds over several levels of experience.

My major conclusion can now be restated: the internal self, synonymous with Freud's "id," is the fount of all consciousness; the external self, synonymous with Freud's "ego," is a learnt representation that is unconscious in itself, but can be consciously "thought with" when cathected by the id; the abstracted self, which provides the reflexive scaffolding for the "superego," is likewise unconscious, but it can consciously "think about" the ego. Because the ego stabilizes the consciousness generated in the id, by transforming a portion of affect into conscious perception—mental solids (and into consciousness *about* perceptions: verbal re-representations)—we ordinarily *think of our selves* as being conscious.

[16] According to the theory of narcissism, this process of separation results initially in a phantasized split between an introjected "me" and a projected "not-me," grounded in the pleasure–unpleasure distinction rather than in the reality principle (Freud, 1925b). Hence Freud's famous dictum to the effect that "hate, as a relation to objects, is older than love" (1915a, p. 139). The projected "bad" object forms the nucleus of the later superego. But this object (which Melanie Klein called the "primitive superego") is a second-person representation. The third-person perspective, which finally enables the self to re-represent itself objectively, *from the viewpoint of the object*, paves the way for the formation of the superego proper.

This obscures the fact that we simply *are* conscious, and our conscious thinking (and perceiving, which thinking represents) is *constantly accompanied by affect.* This constant "presence" of feeling is the background *subject* of all cognition, without which consciousness of perception and cognition *could not exist*. The primary subject of consciousness is literally invisible, so we first have to translate it into perceptual–verbal imagery before we can "declare" its existence. Small wonder, therefore, that it is so regularly overlooked. But the id is only dumb in the glossopharyngeal sense. In truth, it constitutes the primary stuff from which minds are made; we therefore ignore it at our peril. As Freud once remarked, in an almost opposite context:

> The property of being conscious or not is in the last resort our one beacon-light in the darkness of depth-psychology. [Freud, 1923, p. 18]

Later, when he was confronted by the behaviorist juggernaut that was about to sweep aside his life's work, Freud remarked that consciousness was:

> . . . a fact without parallel, which defies all explanation or description. Nevertheless, if anyone speaks of consciousness we know immediately and from our most personal experience what is meant by it. . . . One extreme line of thought, exemplified in the American doctrine of behaviourism, thinks it possible to construct a psychology which disregards this fundamental fact! [Freud, 1940, pp. 157, 157fn.]

We are thus led full circle. To re-establish the difference between behaviorism and psychoanalysis—the science of the mental *subject*—more than a century after Freud first introduced the notion of an unconscious mind (the validity of which is accepted more widely today than it ever was before), we must embrace *consciousness* once more as being the most fundamental feature of the mental.

10 If the id is conscious . . .

The realization that Freud's id is intrinsically conscious has massive implications for psychoanalysis. In this article I can only make a first approach toward the enormous theoretical task that now lies before us if we are going to fully comprehend these implications. I shall draw attention to just four problematical questions in Freudian metapsychology that this revision begins to resolve.

(1) I have explained how perceptual representations attract consciousness by dint of their salience, and how this fits with Freud's view that the most primitive form of cognition (primary-process wishful cathexis) entails hallucinatory wish-fulfillment. It is definitional of hallucinatory processes that they should be conscious. But such wishful phantasies are said to form the nucleus of the system unconscious. This can only mean that the system unconscious revolves around a network of *repressed* hallucinatory phantasies. I am surprised that more commentators have not noticed that this implies that "the" unconscious is hived off from perceptual and cognitive processes, that it is derivative from initially *conscious* experiences and from learning. To my knowledge, only Barry Opatow (1997) has recognized this contradiction, which implies that an innate system pre/conscious *precedes* the development of the system unconscious in mental maturation. Small wonder that Freud was obliged to introduce the "id" concept, in which the representational "system unconscious" was reduced to a mere portion of id called "the repressed."

(2) But if the id is conscious, then what does the repressed consist in? If we retain Freud's view that repression concerns representational processes, it seems reasonable to suggest that repression must involve withdrawal of *declarative* consciousness. This has the effect of reducing an "episodic" cognitive process to an "associative" one (procedural or emotional). The subject of repression still activates the object-presentations in question, but the associative links between them (the "object relations") no longer attract representational–reflexive awareness. We recall that this was the original purpose of ego development: the goal of all learning is *automatization* of mental processes—that is, increased predictability and reduced surprise. It is the biological salience of prediction errors—mediated by attention—that requires the affective "presence" of the id. As soon as the ego has mastered a mental task, therefore, the relevant associative algorithm is automatized. This could be the mechanism of repression: it could consist in a *premature* withdrawal of reflexive awareness (of episodic "presence"), *premature automatization* of a behavioral algorithm, before it fits the bill. In this context, fitting the bill implies obeying the reality principle. Premature automatization therefore results in constant prediction-error, with associated release of free energy (affect), and the ongoing risk of the repressed cognitive material reawakening attention. This lays the foundations for a "return of the repressed," the classical mechanism of neurosis. The therapeutic task of psychoanalysis, then, would still be to undo repressions (to allow the associative links to regain episodic status),[17] in order to enable the reflexive subject to properly master the object relations they represent and generate executive programs more adequate to the task, so that they may then be legitimately automatized. This formulation resolves the awkward disjunction between the so-called cognitive and Freudian unconscious.

(3) Embedded within the many statements that Freud made to the effect that consciousness was a cortical function, by which he seemed mainly to mean a "declarative" function, he always acknowledged the exceptional role of affect. For example:

> It dawns upon us like a new discovery that only something that has once been a Cs. perception can become conscious, and that anything arising from within (*apart from feelings*) that seeks to become conscious must try to transform itself into external perceptions: this becomes possible by means of memory-traces. [Freud, 1923, p. 20; emphasis added]

In other words, although Freud thought that affects were (interoceptive) cortical perceptions, he always recognized that they were felt *directly*. He did not share the view that affects first needed to be exteroceptively represented, or cognitively labeled in working memory, to exist. In fact, for Freud, affects *could not* be represented in the same way that external objects were. This set them apart from all cognitive processes:

> It is surely of the essence of an emotion that we should be aware of it, i.e., that it should become known to consciousness. Thus the possibility of the attribute of unconsciousness would be *completely excluded* as far as emotions, feelings and affects are concerned. [Freud, 1915b, p. 177; emphasis added]

I hope that the neuroscientific facts reviewed here will help us to make better sense of this observation, which, to Freud's credit, he always acknowledged, notwithstanding the theoretical difficulties it must have caused him.

(4) I have already quoted Freud's claim to the effect that "feelings of pleasure–unpleasure govern the passage of events in the id with despotic force . . . the id obeys the inexorable pleasure principle" (Freud, 1940, p. 198). But how can the id be governed by the pleasure principle if it is unconscious in itself, if it is devoid of consciousness, if feelings of pleasure–unpleasure are actually generated on the cortical surface of the ego? If affective consciousness was generated cortically, *the*

[17] Cf. the process of "reconsolidation" mentioned in footnote 10.

pleasure principle would entail top-down control of the id by the ego, which obviously cannot be correct. The primacy of the pleasure principle is therefore affirmed by our relocation of consciousness to the id, and so is the *inhibitory* nature of the ego's top-down influence.

11. The deepest insight

I will close with an aspect of Freud's successive models of the mind that was more essential than the locus and extent of consciousness: namely, his fundamentally *dynamic* conception of it, coupled with the dimension of *depth* (or hierarchy) in the mind. This is why Freud repeatedly stated that the best insight he ever had was that there are two different states of mental energy: one in which cathexis is tonically bound, used for thinking (potential action) rather than action proper, and the other in which it is freely mobile and presses for discharge:

> In my opinion this distinction represents the deepest insight we have gained up to the present into the nature of nervous energy, and I do not see how we can avoid making it. [Freud, 1915b, p. 188]

This fundamental distinction is not only preserved in my proposed revision of Freud's model, along with much else, but it is actually enhanced. The link between affectivity on the one hand and Helmholtz's "free energy" on the other seems to identify a red thread through Freud's work, linking him backward to Helmholtz and forward (via Feynman) to Friston. Considering this and the many other vistas opening up with the rediscovery of the embodied, instinctual brain—which must of necessity be constrained by the cognitive brain, with its predictive modeling—it is difficult to imagine how the neuroscience of the future can be anything but psychodynamic. We are truly living through a Golden Age in neuroscience. As the cognitive neuroscience of the late twentieth century is being supplemented by the affective neuroscience of the present, we are breaking through to a truly *mental* science, and we are finally understanding that the brain is not only an information-processing object, but also an intentional subject.

Still I will end with a whimper rather than a bang. Neuroscience is no more the final court of appeal for psychoanalysis than psychoanalysis is for neuroscience. The final court of appeal for psychoanalysts is the *clinical* situation. Readers are therefore invited to check the theoretical innovations I have introduced here against the data of their psychoanalytic experience. Do these new concepts really make better sense of the facts we observe? Is it really necessary for us to take these difficult steps in our theory?

REFERENCES

Berridge, K. (1996). Food reward: Brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, 20: 1–25.

Blom, J., & Sommer, I. (Eds.) (2012). *Hallucinations: Research and Practice*. New York: Springer.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, 391: 756.

Carhart-Harris, R., & Friston, K. (2010). The default mode, ego functions and free energy: A neurobiological account of Freudian ideas. *Brain*, 133: 1265–1283.

Craig, A.D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, 10: 59–70.

Damasio, A. (1999a). Commentary to Panksepp, J.: Emotions as viewed by psychoanalysis and neuroscience. *Neuropsychoanalysis*, 1: 38–39.

Damasio, A. (1999b). *The Feeling Of What Happens: Body, Emotion and the Making of Consciousness*. New York: Harcourt Brace.

Damasio, A. (2010). *Self Comes to Mind.* New York: Pantheon.

Damasio, A., Damasio, H., & Tranel, D. (2012). Persistence of feeling and sentience after bilateral damage of the insula. *Cerebral Cortex*. 23 (4): 833–846.

Ebisch, S., Salone, A., Ferri, F., de Berardis, D., Romani, G., Ferro, F., et al. (2012). Out of touch with reality? Social perception in first-episode schizophrenia. *Social Cognitive & Affective Neuroscience*. 8 (4): 394–403.

Freud, S. (1891). *On Aphasia*. *Standard Edition*, 14: 209–215.

Freud, S. (1894). The neuro-psychoses of defence. *Standard Edition*, 3: 45–61.

Freud, S. (1895). Project for a scientific psychology. *Standard Edition*, 1: 281–397.

Freud, S. (1896). *Extracts from the Fliess Papers* [Letter of 1 January]. *Standard Edition*, 1: 388–391.

Freud, S. (1900). *The Interpretation of Dreams. Standard Edition*, 4/5.

Freud, S. (1911). Formulations on the two principles of mental functioning. *Standard Edition*, 12: 215–226.

Freud, S. (1915a). Instincts and their vicissitudes. *Standard Edition*, 14: 117–140.

Freud, S. (1915b). The unconscious. *Standard Edition*, 14: 166–204.

Freud, S. (1916–17). *Introductory Lectures on Psycho-Analysis. Standard Edition*, 15/16.

Freud, S. (1917). Metapsychological supplement to the theory of dreams. *Standard Edition*, 14: 222–235.

Freud, S. (1920). *Beyond the Pleasure Principle*. *Standard Edition*, 18: 7–64.

Freud, S. (1923). *The Ego and the Id*. *Standard Edition*, 19: 12–59.

Freud, S. (1925a). A note upon "the mystic writing-pad". *Standard Edition*, 16: 227–232.

Freud, S. (1925b). Negation. *Standard Edition*, 19: 235–239.
Freud, S. (1940). *An Outline of Psychoanalysis. Standard Edition*, 23: 144–207.
Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, 11: 127–138.
Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, (62) (2): 1230–1233.
Gallese, V. (2011). *Bodily Selves in Relation: Embodied Simulation and Intersubjectivity.* Paper presented at the 12th Annual International Neuropsychoanalysis Congress, Berlin, 25 June.
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119: 593–609.
Gloor, P. (1992). Role of the amygdale in temporal lobe epilepsy. In: *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, ed. J. Aggleton. New York: Wiley-Liss, pp. 505–538.
Huston, J., & Borbely, A. (1974). The thalamic rat: General behaviour, operant learning with rewarding hypothalamic stimulation, and effects of amphetamine. *Physiology & Behavior*, 12: 433–448.
Kihlstrom, J. (1996). Perception without awareness of what is perceived, learning without awareness of what is learned. In: *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*, ed. M. Velmans. London: Routledge, pp. 23–46.
James W. (1890). *The Principles of Psychology.* New York: Henry Holt.
Lange, C.G. (1885). *On Emotions: A Psycho-Physiological Study*. Baltimore, MD: Williams and Wilkins.
LeDoux, J. (1999). Psychoanalytic theory: Clues from the brain. *Neuropsychoanalyis*, 1: 44–49.
Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences,* 8: 79–86.
Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral Brain Sciences*, 30: 63–134.
Mesulam, M. M. (2000). Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system and hemispheric lateralization. In: *Principles of Behavioral and Cognitive Neurology* (2nd edition). New York: Oxford University Press, pp. 1–120.
Moruzzi, G., & Magoun, H. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, 1: 455–473.
Opatow, B. (1997). The real unconscious: Psychoanalysis as a theory of consciousness*. Journal of the American Psychoanalytic Association*, 45: 865–890.
Panksepp, J. (1998). *Affective Neuroscience*. New York: Oxford University Press.
Penfield, W., & Jasper, H. (1954). *Epilepsy and the Functional Anatomy of the Human Brain*. Oxford: Little & Brown.
Petkova, A., & Ehrsson, H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS ONE*, 3: e3832.
Pfaff, D. (2006). *Brain Arousal and Information Theory.* Cambridge, MA: Harvard University Press.
Shewmon, D., Holmse, D., & Byrne, P. (1999). Consciousness in congenitally decorticate children: Developmental vegetative state as a self-fulfilling prophecy. *Developmental Medicine and Child Neurology*, 41: 364–374.
Solms, M. (1997). What is consciousness? *Journal of the American Psychoanalytic Association*, , 45: 681–778.
Solms, M. & Panksepp, J. (2012). The id knows more than the ego admits. *Brain Sciences*, 2: 147–175.
Strachey, J. (1962). The emergence of Freud's fundamental hypotheses. In: S. Freud, *Standard Edition*, 3: 62–68.
Sur, M., & Rubenstein, J. (2005). Patterning and plasticity of the cerebral cortex. *Science*, 310: 805–810.
Tsakaris, M. (2011). *The Neurophilosophy of Embodied Cognition and Agency.* Paper presented at 12th Annual International Neuropsychoanalysis Congress, Berlin, 23 June.
Wright, J., & Panksepp, J. (2012). An evolutionary framework to understand foraging, wanting, and desire: The neuropsychology of the SEEKING system. *Neuropsychoanalysis*, 14 (1): 5–39.

Repression as the Condition for Consciousness

Commentary by Ariane Bazan (Brussels)

Mark Solms makes a convincing case for the subcortical structures as consciousness-inducing instances, and the audacious logical consequence that therefore the Freudian id is conscious makes sense. However, in his understanding of this conscious id, affect and drive are conceptually confused and a brain-based view of affect, drive, and pleasure altogether is defended. My first aim is to stress the importance of understanding the drive—the vector between an internal body regulatory imbalance and an external body motor response—as an acquired link, for which the criterion is first given by pleasure, produced by a release of tension when an internal body need is alleviated. Moreover, I question the representational nature of this primary subjective consciousness, and I propose that the constitutive contribution of the neocortex to consciousness is not so much memory space than the process of inhibition—or repression. This enables the distinction between a mental and a perceived object and consequent action-selection and, in the process of doing so, generates representations and "solidifies" objects.

Keywords: affect; drive; inhibition; pleasure; representation; repression

Mark Solms's Target Article has the great merit of putting a range of empirical observations together and pointing out the "obvious" (which nobody has done before)—that "cortical varieties of consciousness actually depend on the integrity of subcortical structures, not the other way round" (Section 5). Moreover, it makes sense to me to understand these subcortical structures, and the consciousness they generate, as of an inherent drive-nature and therefore all the more closer to the concept of the Freudian "id."

My first aim is to defend a conceptual distinction between an affect and a drive (or "instinct"), in the Solms–Panksepp notion of the conscious "id," whereas both concepts are often equated or conflated in the article.[1] I understand that one of the pillars of the conceptual framework presented by Solms is Panksepp's (1998) ground-breaking notion of affective neuroscience, implying that there are distinct subcortical brain circuits, which seem conserved over species and which, when activated, give rise to the behavioral expressions of the different affects and therefore, as the result of the feeling of these motor discharges, to the different emotions. Panksepp (1998) regards these emotional operating systems not as drives but as regulatory mechanisms emerging from the intrinsic potentials of the nervous system. Solms also considers the brain as the emergence site of affect ("Affect is an intrinsic property of the brain"), of the affect-as-the-drive ("They [the various instinctual motivational circuits] . . . are intrinsic to the brain"), and of the (un-) pleasure feelings ("Feelings of pleasure–unpleasure … are readily generated by stimulating a region of the ERTAS knows as periaqueductal grey"). This suggests an all-brain organization for affect, drive and pleasure. Therefore, it is as if an organism comes to organize itself as an intentional entity as the result of a causal chain emerging from the brain, from its "intrinsic potentials," and, as a consequence, eventually from the innate structure of the brain.[2] Flexibility and learning are understood as mere ways to fine-tune the system to changing environments, but the major impulse comes from the nervous system, be it from the subcortical and brainstem structures.

Now, it is difficult to understand intentionality itself as emerging from the brain. My fundamental point is that *we cannot do away with the body as a site of origin*—in contrast to a brain-based site of origin—if we want to end up with directed, oriented behavior—that is, with an intentional system. Going back to Freud's (1895) concept of the "experience of satisfaction," a possible chain of event might go as follows. The main regulatory systems (respiration, digestion, sudation, excretion, circulation, copulation, etc.) are taken care of by the viscera of the internal body. Suppose that there is a depletion of nutrients in the tissues of the internal body; this lack is translated in an afferent stimulation, which is "achieved through a network of upper-brainstem, diencephalic, and basal forebrain 'arousal' struc-

[1] For example, "affect and affective motivation"; "various instinctual motivational circuits … known as the circuits for 'basic emotions'"; "What distinguishes them [the basic emotions] is their *instinctual* nature"; etc. See also Shevrin (2003) for a comparison between Panksepp's SEEKING system and Freud's definition of the drive.

Ariane Bazan: Université Libre de Bruxelles, Brussels, Belgium.

[2] For example, "Freud . . . recognized that the basic emotions are innate mental organizations (cf. Panksepp, 1998)." This kind of statement is confusing in the context of this article, since this might be true for the basic emotions (a motor pattern set off by a stimulation, internal or external; the link is innate or learned; what is innate is the activated pattern of reactions), but this seems to me fundamentally wrong when it comes to drive (which by definition is a vector linking an internal body need state to an external motor behavior pattern; it is of crucial importance to see that the linking is not innate but acquired; see discussion below).

tures . . . known . . . as the extended reticulo-thalamic activating system." This arousal stimulation will probably set off the SEEKING system. Panksepp (1998, p. 194) indicates that "The species-typical expressions of this system lead to foraging in some species and predatory stalking in others." But even in animals, only in the case where foraging or predator stalking leads to food coming in will the depletion, from which the urge to act first emerged, get replenished. This replenishing is a release of tension, which Freud (1895) qualifies as "pleasure." In other words, *the pleasure criterion is given by the (internal) body*—not by the brain. Only if the motor pattern chosen is successful will there be pleasure, a pleasure given by (the release of tension in) the body.

This pleasure has to be distinguished from another kind of gratification. The behavior that has delivered this pleasure—because it has delivered this pleasure—will get tagged physiologically as a salient behavior. This is what neuroscience tells us (e.g., Berridge, 2009) but also what Freud (1905, p. 182) suggested: "This satisfaction must have been previously experienced in order to have left behind a need for its repetition; and we may expect that Nature will have made safe provisions so that this experience of satisfaction shall not be left to chance." As a result, once this tagging happened, acting out the behavior delivers a gratification subserved by the dopamine circuitry.[3] Say that due to changed circumstances (e.g., captivity) some other motor pattern (e.g., pushing a button) leads to food coming in; then it is that motor pattern that will acquire an incentive value. In other words, the link between the internal body's need status and the external body's behavior pattern responding to that need has to be physiologically registered, by the internal body's signal of pleasure when the need is effectively alleviated; even in animals, therefore, this link is acquired and not innate.

[3] Note that this distinction between the pleasure given by the consumption of something that the internal body was in need of and the *gratification* or the *incentive* given by the motor activation of the behavior pattern leading to this consumption parallels Berridge's (2009) distinction between "liking" and "wanting" (see also Shevrin, 2003) and the Lacanian distinction between pleasure and *jouissance* (Lacan, 1959–1960; see also Bazan & Detandt, in press; Bazan, Detant, & Tiete, 2012). Also note that the motor behavior pattern may remain physiologically gratifying, because it is carved as such in the physiology of the subject, even when, due to changed circumstances, it no longer brings pleasure. This disconnection between pleasure and gratification happens far more in a human than in animal life, due to the large "translation gap" between the internal and external body, which induces a large variability in the range of possible adequate actions, many of which may later on in life become inadequate, while still being gratifying. This may then result in the typical suffering that is induced by the persistence of behaviors that the subjects themselves do not find pleasurable—that is, it is one of the major causes of human distress. See also Johnson (2008).

But there is more. In most animals, the "translation gap" between internal and external body is small: there is almost always immediate efficient adjustment between the internal regulatory systems and the external behaviors, which are even qualified in the article as "associated peremptory" actions. However, in humans, due to the large helplessness of the newborn, that translation gap becomes an unavoidable reality. For example, the kind of behavior induced by the human SEEKING system in a newborn when activated by an internal body need signal is far less specific than the complex stereotyped behavior of newborn animals: when it comes to "seeking," the human baby will cry and wriggle. Nonetheless, this behavior may be effective. Indeed, a mother may come along, hear the baby cry, *interpret* the behavior, and feed the baby. In the end, as the milk alleviates the hunger and thereby brings pleasure, the cry was an adequate act and gets physiologically tagged as such—that is, the motor activity of crying and sucking become gratifying, which ensures that they will in the future be readily repeated when a new hunger signal arises. In other words, humans, far more than animals, need the internal body tissues-based pleasure criterion to connect afferent brainstem stimulation with "associated" motor expression patterns, which are much less a-priori associated (as might confusingly be understood from, for example, "[instincts] are intrinsic emotional stereotypes").

For these reasons, we cannot do away with the body as a site of origin for the constitution of an organism, and eventually of a mental system. As the notion of "drive" presupposes this bodily site of origin in its definition, while an affect may also arise from the encounter with an external stimulus, it is important not to conflate affects and drives. It is important to remember that Freud (1915a) called the drive the vector between the biological and the mental.

Second, what kind of consciousness is generated by the subcortical structures? Solms indicates that in the Libet experiment, the "primary subjective" or the "affective (primary)" self actually initiates the movement and that it is only to the "representational 'declarative'" self that awareness of the movement comes with some delay. This makes sense, but I don't think that this primary subjective self is the self that generates the kind of consciousness that gives us the (illusory) idea of free will, and, therefore, the Libet results remain: the awareness of having decided to move comes after the decision to move. What is more, Haggard and Eimer (1999), for example, show that this awareness may come as the result of the decision to move and, more precisely, as the result of the movement *selection*—that is, of the inhibition of nonchosen alternatives.

Let us go back to the experience of satisfaction. Say the hungry baby cried and the mother fed him. The whole motor pathway leading to the effective sucking of the breast, the adequate act, gets linked to the hunger drive and becomes physiologically gratifying. In particular, the image of the breast in the right angle for effective sucking becomes the wishful image, which serves as the reference for the motor search of the head movements. In the first instance, as Solms keenly points out, whenever the hunger stimulus appears, it is that motor image that gets activated and there is a hallucinatory wish-fulfillment with a release of the sucking movement. Indeed, in this stage one could say that "biologically valenced . . . objects of past experience are rendered conscious by dint of their 'incentive salience'". I agree that this kind of consciousness, characterized by hallucination and acting-out, is played-out through neocortical activation aroused by drive-instigated subcortical activation and that it does not entail secondary-process cognition. But I think that this "very primitive form" of disinhibited hallucinatory consciousness might not be representational, and it might even be open to debate whether this state then really qualifies as "conscious" if it is not representational.[4]

I propose, however, that the state we would readily qualify as conscious emerges when no breast is present. In that case, releasing the sucking movement not only will not be effective, but will also lead to a loss of energy. The baby would be better off to stop sucking and start crying again. Therefore, it becomes crucial for the baby's survival to be able to distinguish a mental image from a perceived image of a breast. It is there where the ω neurons—the motor neurons of perception—come in (see also footnote 6 in the Target Article). As soon as there is enough inhibitory weight of the maturing ego,[5] the ω neurons (e.g., oculomotor neurons)—and especially their "messages of discharge" or "indication of reality"[6] (Freud (1895, p. 325)—will enable this distinction, because a movement of the eyes has radically more drastic effects on a perceived breast than on a mental image of a breast. At this stage, as the internal origin of the mental image is recognized, the sucking movement is withheld. We might say that the sucking action "is not hypercathected, remains thereafter in the Ucs." (Freud, 1915b, p. 202)—that is, we have here a very basic form of repression (namely, of the motor act of sucking).[7]

At the same time, the neuroscientist Jeannerod (1994, p. 201) suggests that, if no breast is present, the neurons encoding the final configuration

> . . . (of the environment, of the body, of the moving segments, etc.) as they should arise at the end of the action . . . remain active until the requested configuration has been obtained. If the goal [of an action plan] were not reached, the sustained discharge would be interpreted centrally as a pure representational activity and give rise to mental imagery.

In other words, the baby will still generate an internal image of the breast, but this image will be recognized as a mental image and is thus no longer a hallucination. This means that it is truly *inhibition* of action that generates representations. Moreover, what is called an "object" can only be what is assembled in these representations as a wishful action or a desired goal of that action (i.e., as an object of a drive, ultimately).[8]

Furthermore, the kind of consciousness we experience thanks to the contribution of the neocortex is the kind of consciousness that we end up with after *inhibition* has made selection possible, stabilizing both the selected and the inhibited actions/objects. Therefore, I do not think it is so much the passive presence of memory space that the cortex contributes to consciousness, as Solms suggests; I propose, rather, that it is the active process of inhibition (or, in psychodynamic terms, of *repression*; see Bazan, 2012; Bazan and Snodgrass, 2012) that is the foundational, constitutive process that the cortex contributes to representational consciousness. And this kind of consciousness, to my view, is truly secondary-process consciousness. For example, the Swiss psychiatrists Saraga and Gasser (2005, p. 111) indicate that Freud underscored the im-

[4] Is a hallucination a representation, or should we consider it as an activation or stimulation at the periphery of the mental apparatus—namely, at its perceptual periphery—in the same way that acting-out is an activation at the motor periphery of the mental apparatus?

[5] Or of the Default Mode Network (Carhart-Harris & Friston, 2012).

[6] Elsewhere, we have defended an equivalence between the Freudian "indications of reality" and the "efference copies" of the modern neuroscientific models (Bazan, 2007a; Bazan & Snodgrass, 2012; see also Shevrin, 1998).

[7] Indeed, the French psychoanalyst Le Guen (2001, p. 46) underscores that "what has to be inhibited in fact not the object, but truly the motor act, as a function."

[8] When does an object become an object to a mental apparatus or a Freudian object-presentation, as Solms points out? There is only one criterion possible *from within* the emerging mental system: when it becomes a possible object of the drive. This is also what Solms says when he speaks of "objects of desire coming to mind." By trial and error, by learning, through interpretation from others, internal body need states get linked to a range of *adequate* actions, delivering the objects that can alleviate the depletion at the origin of the drive. This makes each adequate action and its object an entity. We usually think of objects as perceptual entities, while even in Freud's object-presentation model the motor modality—the usual way of motor interaction with that object, its "grasp"—is present as an important constitutive component. Likewise, the neuroscientists Grabowski , Damasio, and Damasio (1998) and Grafton, Fadiga, Arbib, and Rizzolatti (1997) propose that objects are encoded as the motor program that we have to mobilize to use these objects. In that sense, there is some neurophysiological equivalence between the adequate act and the adequate object of a drive.

portance of this inhibition as being the essence of the secondary process, which enables the development of thought itself, the "substitute of the hallucinatory wish-fulfillment."

A final point on words and things. I disagree with Solms on understanding "word-presentations" as re-representations of objects at a higher level in secondary-process thinking. Freud (1891), with his model in *On aphasia*, intended in the first place to stress a certain *equivalence* between words and objects. Both words and object-presentations are defined by the same type of characteristics—namely, perceptual and motor characteristics for both. The perceptual characteristics for objects mostly cover the whole spectrum (vision, smell, taste, feeling, sound, etc.), whereas for words the range is more restrained (the graphic image, the word sound). The motor characteristics of objects include the usual way we interact with them, while for words it includes the articulation movement for the spoken word and the writing movement for written words. In other words, a word is, in the first place, an object like another and is treated as another. The special faculty that emerges from language only emerges because of the fact that connections are made between specific object-representations and specific word-representations (the famous double link in his scheme) at a conscious level. But this linking—the reference capacity of language—is structurally unstable, very much so in the unconscious (where the word-presentation is loosely connected to the object-presentation) but also consciously, due to language's structural ambiguity. So, even if in some cases word-presentations are re-representations of objects at a higher level, they always also are not—that is, they can induce effects without any connection to their "corresponding" object-presentations[9] (such as in signifier-structured symptoms; Bazan, 2007b, 2011b). It is precisely because words can navigate on these two hierarchical levels—the primary process and the secondary process—that they are effective in the talking cure, and not only because they are the instruments of rational, contextualized thought capable of representing relations.

Finally, there are three things with which Solms closes his paper and with which I wholeheartedly agree. First, I completely agree that "it is difficult to imagine how the neuroscience of the future can be anything but psychodynamic."[10] Second, when Solms says that "we are breaking through to a truly *mental* science," I completely adhere to this view (and I have defended elsewhere that this will be the logical result of the tremendous revolution in the neurosciences; Bazan, 2011a). Let me add, along the same line of thought, that I also adhere to his word choice of "mental solids" and that we are reminded that Freud (1900, p. 613) spoke about "psychic reality" and not about some metaphorical discursive construction for his understanding of the unconscious. And, third, I am of course encouraged to read that the final word is given to clinical expertise. As psychoanalysts, and especially in the domain of neuropsychoanalysis, we have not been sufficiently proud and aware over the last decades of our precious and unique clinical method. This lengthy, most of the time undirected, often times confusing and ambiguity-inducing therapeutic offering is so much at the antipodes of what has been valued in science, and in clinics, that many of us have been readily willing to leave it or to undervalue it—whereas it is actually at the very core of our unique contribution.

[9] For empirical evidence on this point see Villa, Shevrin, Snodgrass, Bazan, and Brakel (2006).

[10] "Psychoanalysis is the future of neurosciences" was also the closing sentence of my presentation at the Athens' International Neuropsychoanalysis Congress (Bazan, 2012).

REFERENCES

Bazan, A. (2007a). An attempt towards an integrative comparison of psychoanalytical and sensorimotor control theories of action. In: *Attention and Performance XXII*, ed. P. Haggard, Y. Rossetti, & M. Kawato. New York: Oxford University Press, pp. 319–338.

Bazan, A. (2007b). *Des fantômes dans la voix. Une hypothèse neuropsychanalytique sur la structure de l'inconscient.* Collection Voix Psychanalytiques. Montreal: Editions Liber.

Bazan, A. (2011a). The grand challenge for psychoanalysis—and neuropsychoanalysis: Taking on the game. *Frontiers in Psychology,* 2: 220.

Bazan, A. (2011b). Phantoms in the voice: A neuropsychoanalytic hypothesis on the structure of the unconscious. *Neuropsychoanalysis,* 13 (2): 161–176.

Bazan, A. (2012). From sensorimotor inhibition to Freudian repression: Insights from psychosis applied to neurosis. *Frontiers in Psychology,* 3: 452.

Bazan, A., & Detandt, S. (in press). De la jouissance au wanting. Physiologie d'un concept Lacanien. *Revue Belge de Psychanalyse*.

Bazan, A., Detant, S., & Tiete, J. (2012). *From "Jouissance" to "Wanting": On the Relevance of a Lacanian Concept to Understand the Wiring of the Brain in Addiction.* Poster presented at the 13th International Neuropsychoanalysis Congress, Athens, 14–16 June.

Bazan, A., & Snodgrass, M. (2012). On unconscious inhibition: Instantiating repression in the brain. In: *Trends in Neuropsychoanalysis: Psychology, Psychoanalysis and Cognitive Neuroscience in Dialogue*, ed. A. Fotopoulou, D. W. Pfaff, & E. M. Conway. Oxford: Oxford University Press, pp. 307–337.

Berridge, K. C. (2009) “Liking” and “wanting” food rewards: Brain substrates and roles in eating disorders. *Physiology & Behavior,* 97 (5): 537–550.

Carhart-Harris, R. L., & Friston, K. J. (2012). Free-energy and Freud: An update. In: *Trends in Neuropsychoanalysis: Psychology, Psychoanalysis and Cognitive Neuroscience in Dialogue*, ed. A. Fotopoulou, D. W. Pfaff, & E. M. Conway Oxford: Oxford University Press, pp. 219–229.

Freud, S. (1891). *On Aphasia*, trans. E. Stengel. New York: International Universities Press, 1953.

Freud, S. (1895). Project for a scientific psychology. *Standard Edition,* 1: 281–392.

Freud, S. (1900). *The Interpretation of Dreams*. *Standard Edition,* 4/5.

Freud, S. (1905). *Three Essays on the Theory of Sexuality*. *Standard Edition,* 7.

Freud, S. (1915a). Instincts and their vicissitudes. *Standard Edition,* 14: 117–140.

Freud, S. (1915b). The unconscious. *Standard Edition,* 14: 159–215.

Grabowski, T. J., Damasio, H., & Damasio, A. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *NeuroImage,* 7: 232–243.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *NeuroImage,* 6: 231–236.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research,* 126: 128–133.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences,* 17: 187–245.

Johnson, B. (2008). Just what lies “beyond the pleasure principle”? *Neuropsychoanalysis*, 10: 201–2012.

Lacan, J. (1959–1960). *Le séminaire. Livre VII. L’éthique de la psychanalyse, 1959–1960 (Le Champ freudien)*. Paris: Seuil, 1986.

Le Guen, C. (2001). “Quelque chose manqué . . .” De la répression aux représentations motrices. *Revue Française de Psychanalyse* 65: 37–70.

Panksepp, J. (1998) *Affective Neuroscience*. New York: Oxford University Press.

Saraga, M., & Gasser, J. (2005). Epreuve de réalité et psychose chez Freud. La fin de la psychose à l’heure d’un dernier “retour”? *Psychothérapies,* 2 (25): 109–115.

Shevrin, H. (1998). Why do we need to be conscious? A psychoanalytic answer. In: *Advanced Personality*, ed. D. F. Barone, M. Hersen, & V. B. New York: Plenum Press.

Shevrin, H. (2003). *The Psychoanalytic Theory of Drive in the Light of Recent Neuroscience Findings and Theories*. Paper present at the 1st Annual C. Philip Wilson, MD Memorial Lecture, New York, 15 September.

Villa, K. K., Shevrin, H., Snodgrass, M., Bazan, A., & Brakel, L. A. W. (2006). Testing Freud’s hypothesis that word forms and word meanings are functionally distinct in the unconscious: Subliminal primary process cognition and its links to personality. *Neuropsychoanalysis,* 8: 117–138.

The Brainstem Begs the Question: *"Petitio Principii"*

Commentary by Heather Berlin

Mark Solms proposes that the upper brainstem is intrinsically conscious and that the cortex is intrinsically unconscious and is only permeated with consciousness from the brainstem. His theory relies heavily on studies of hydranencephalic children, who appear to have emotional reactions to outside stimuli despite the fact that they are missing a cerebral cortex. Solms uses this as his main evidence that consciousness is not a function of the cortex. However, I explain in this commentary, based on years of accumulated neuroscientific evidence, why Solms is making two unsupported assumptions: (1) that without cortex you have affective consciousness, and (2) that without brainstem you lose consciousness. It is important not to confuse "consciousness as such" (i.e., wakefulness) with the "content of consciousness" (i.e., awareness). There is excellent converging evidence for the cortical basis of the *contents* of consciousness.

Keywords: consciousness; cortico-thalamo-cortical pathways; frontoparietal network; hydranencephaly; neuroscience; persistent vegetative state

In his eloquent article "The Conscious Id," Mark Solms proposes several groundbreaking ideas, which, if substantiated, could potentially turn the fields of neuroscience and psychoanalysis on their heads. He first suggests that "affective consciousness" is derived from brainstem mechanisms that control and receive input from the autonomic body, and that "cognitive consciousness" is derived from cortical mechanisms that receive and send information to and from the sensorimotor body. This in itself is not so radical, but he goes on to propose that all of our cortically-based sensory and perceptual experiences are imbued with consciousness only by the affective processes that exist to govern our internal bodily needs. Solms therefore makes the radical claim that consciousness is a function of the upper brainstem. He proposes that the upper brainstem is intrinsically conscious and the cortex is intrinsically unconscious and is only permeated with consciousness from the brainstem. I have several major points of contention with this proposal, based on the accumulation of years of neuroscientific evidence.

First and foremost, Solms's theory relies heavily on one piece of evidence: Bjorn Merker's 2007 study of hydranencephalic children—that is, children born without a cortex. In Merker's study, hydranencephalic children appear to have emotional reactions to outside stimuli despite the fact that they are missing a cerebral cortex. Although they have no cortex, Solms reports that they clearly display signs of feeling pleasure and unpleasure and an extensive capacity for emotional learning. But emotional learning and processing rewards and punishments does not require consciousness (Berlin, 2011; Esteves, Parra, Dimberg, & Ohman, 1994; Fischman, 1989; Lamb et al., 1991; Pessiglione et al., 2007). He uses this as his main evidence that consciousness is not a function of the cortex—that is, that you can have consciousness in the absence of a cortex. However, we cannot assume that expressions of emotion equate with consciousness, when they may just be reflexive. Changes in vigilance and expression of emotion do not equal consciousness—consider, for example, persistent vegetative state (PVS) patients. Solms makes two unsupported assumptions: (1) that without cortex you have affective consciousness, and (2) that without brainstem you lose consciousness.

Early behaviors are highly reflexive, and specific arousal effects, as Pfaff (2006) has catalogued, are most of a lower organism's behavioral repertoire, but whether there is a conscious phenomenal aspect of these phenomena is unknown. The use of the hydranencephalic infant model (Merker, 2007; Shewmon, Holmse, & Byrne, 1999) begs the question entirely: showing that strongly conserved emotional facial displays and conditioned responses from a brainstem/spinal-cord system can be developed over time says nothing about whether conscious emotional states attach to these observable phenomena.

Therefore, Solms's primary assumption that hydranencephalic children are conscious is unwarranted. We cannot assume that having a sleep–wake cycle and expressions of emotion (laughter, rage, etc.) necessitates consciousness. For example, we can reproduce similar pseudo-emotional reactions in nonconscious machines (e.g., affective computing—such as the Siri application on Apple devices). Affective behaviors do not equate with consciousness—for example, decorticate rats, sleepwalkers, and people with conversion disorders and hysterical blindness can all display what look like meaningful affective behaviors without being

Heather Berlin: Department of Psychiatry, Mount Sinai School of Medicine, New York, U.S.A.

conscious of them. While it is true that they may in fact be consciousness, we cannot assume that they are. Unconscious processes can be quite sophisticated and complex (Berlin, 2011).

The crux of Solms's theory relies on a projection of the existence of consciousness based on what look like meaningful emotional behaviors, an example of the "moralistic fallacy" (arguing that something must be true because it would make us feel good to believe it). Humans have a natural desire to assume that consciousness exists (Shermer, 2011), but it is a misconception to uncritically equate vigilance, eye movements, and expression of emotion with consciousness. For example, PVS patients can open and close their eyes, but some do not have sleep–wake cycles as shown with EEG (Landsness et al., 2011). Take, for example, the well-known case of the PVS patient Terri Schiavo. Video clips of Terri showed spontaneous movements and reflexes (e.g., grimacing, crying, eye tracking for brief periods), but no evidence of awareness. In fact, according to the medical examiner's report, she was cortically blind. We must at least consider the possibility that the "emotional behaviors" displayed by hydranencephalic children are simply reflexes. Reflexes can be mediated without consciousness (e.g., classic spinal reflexes in frogs to pain; the withdrawal reflex can be accompanied by consciousness if the cortex is intact). We could just as easily use enteric nervous-system reactions to measure consciousness. Behavioral responses of the enteric nervous system to external stimuli would probably be accurate and reliable, but that does not mean that the enteric nervous system is conscious. What we really need is a theory of consciousness that will enable us to quantify consciousness with an objective, independent measure.

Furthermore, subcortical mediation of consciousness has been described so far only in congenital brain malformations, so developmental plasticity may play a role. Hydranencephalic children's abilities may reflect "vertical" plasticity of brainstem and diencephalic structures. "Vertical plasticity" is subcortical plasticity for supposedly cortical functions, whereas "horizontal" plasticity is cortical plasticity for cortical functions or subcortical plasticity for subcortical functions (Shewmon, Holmse, & Byrne, 1999). In fully formed adult brains, losing cortical function results in loss of the *content* of consciousness. Discrete cortical lesions give rise to specific pathologies of consciousness, such as blindsight, neglect, amnesia, anosognosia, and changes in personality and emotion (e.g., Phineas Gage; Harlow, 1848).

A distinction must be made between "consciousness as such" (i.e., wakefulness) and the "content of consciousness" (i.e., awareness). Enabling factors are necessary for any form of "consciousness as such" (wakefulness) to occur. These enabling factors include the mesencephalic reticular formation (a.k.a. ascending reticular activating system), cholinergic pathways from the brainstem and basal forebrain, and the intralaminar nucleus of the thalamus (Koch, 2004). One could perhaps think of these enabling factors as the power supply to the brain, as distinct from its processing center. However, specific factors are required for any one particular conscious percept—that is, "content." Experiments show that various cognitive tasks that require awareness are accompanied by short-term temporal correlations among distributed populations of neurons in the thalamocortical system. Hence, we need enabling factors as well as a dominant neuronal coalition in the cortex and thalamus for consciousness to occur.

There is no reason to believe that loss of the upper brainstem alone produces permanent unconsciousness unless the lesions are extensive, bilateral, and extend rostrally—and even in these cases the contribution of functional alteration of the rest of the cerebrum is unclear (N. Schiff, personal communication). It is likely that an intact corticothalamic system could in fact recover consciousness without the brainstem. For example, studies in cats show that brainstem lesions can decrease activation, but if you wait long enough their vigilance can recover; they can eventually recover activation and deactivation patterns and slow-wave sleep patterns. Studies by Villablanca (2004) show that you can cut and isolate thalamus and cortex from the brainstem of cats and keep them alive. At first this induces a coma, but after about a month the cortex reactivates and they show sleep–wake cycles. So animals with brainstem lesions can come out of a coma, but if they have no cortex activation they cannot. A cortex without a brainstem can potentially become conscious.

People in a PVS are "awake" (presence of sleep–wake cycles or eyes opening and closing), but not "aware" (no evidence of awareness of self or environment, and an inability to interact with others). The brainstem is mostly spared while the grey and white matter of both cerebral hemispheres are widely and severely damaged. Overall cortical metabolism is about 40–50% of the normal range (Laureys, 2005; Laureys, Lemaire, Maquet, Phillips, & Franck, 1999; Laureys et al., 1999; Schiff et al., 2002). PVS patients usually have either diffuse cortex or thalamic lesions. However, lesions in the brainstem of PVS patients can resolve, but lesions in cortex or paramedical thalamus cause loss of consciousness (Schiff, 2004, 2008). Upper-brainstem lesions can lead to coma, but patients can recover as long as cortex and thalamus are intact (Schiff, 2004, 2008).

While it is true that the normal conscious state depends intimately on the activity of the upper brainstem/central thalamus and related centrencephalic components, lesions that produce initial coma give way to varying patterns of recovery that emphasize difference in the contribution of these structures and opportunities for the conscious state to reconstitute, despite lesions that initially wipe it out. Several papers by Nicholas Schiff address the misconception that the lesion literature localizes consciousness to the upper brainstem and/or central thalamus (Schiff, 2004, 2008). Lesions restricted to the rostral pons and mesencephalon producing coma have roughly dichotomous outcomes—death due to malignant hypotension and cardiopulmonary dysregulation, or recovery of consciousness typically in about 7 days. In sum, there is currently little neurological evidence for any local area to be absolutely critical for consciousness. Instead, many important hubs can alter the critical dynamic processes needed across the cerebrum (primarily corticothalamic systems) to maintain the awake, intentional conscious state (N. Schiff, personal communication).

Functional imaging in transient dissociations of wakefulness and awareness show decreased blood flow in the frontoparietal network in patients with complex partial seizures, absence seizures, and sleepwalking (Laureys, 2005). Medial posterior cortex (including the precuneus and posterior cingulate cortex) is the most active brain region in healthy controls and patients with locked-in syndrome (fully conscious, but paralyzed and thus not behaviorally responsive). In PVS patients, this same brain area is the least active region; patients in a minimal conscious state show an intermediate metabolism here, less than PVS patients, but more than healthy controls. These brain regions are among the most active in conscious waking and among the least active in altered states of consciousness such as general anesthesia, sleep, hypnotic state, dementia, and Korsakoff's or postanoxic amnesia. So this richly connected association area may be part of the neural network subserving awareness/consciousness (Laureys, Owen, & Schiff, 2004).

Neuroimaging of vegetative state (VS) patients identified brain areas that still show activation during external stimulation, but this activation is limited to subcortical and "low-level" primary cortical areas, disconnected from the frontoparietal network necessary for awareness/consciousness (Laureys, 2005). Electrical stimulation (painful in controls) of 15 VS patients activated midbrain, thalamus, and primary somatosensory cortex (S1), but not higher order areas of the pain matrix (secondary somatosensory areas, insular, posterior parietal, anterior cingulate cortex). Also, activated S1 was isolated from the frontoparietal network thought to be required for consciousness perception (Laureys et al., 2002). Similarly, auditory stimulation in VS patients activated primary auditory cortex, but not higher order multimodal areas from which they were disconnected (Boly et al., 2004; Laureys et al., 2000). The activation in primary cortices in these awake, but unaware, patients confirms Crick and Koch's (1995) early hypothesis (based on visual perception and monkey histological connectivity) that neural activity in primary cortices is necessary, but not sufficient, for awareness/consciousness.

In a recent study, Boly et al. (2012) found that decreased backward corticocortical connectivity from frontal to parietal cortices was associated with loss of consciousness under the anesthetic propofol, but thalamocortical connectivity was not. Thus, corticocortical communication appears to be important in the maintenance of consciousness and propofol seems to directly affect these cortical dynamics. What matters is cortex. When cortex comes back, so does consciousness; everything else is doubtful. In line with this, Velly et al. (2007) took intracranial recordings from subthalamic nuclei (thalamus) and cortex in Parkinson's disorder patients during anesthesia (sevoflurane or propofol). When the thalamus was "asleep," there was low-frequency activity for several minutes before the patients became unconscious. Patients only became unconscious when cortex started showing slow waves. Patients remained conscious as long as their cortex was activated, which suggests that consciousness mainly involves the cortex and we may not even need thalamus activation for consciousness.

Summary

There is excellent converging evidence for the cortical basis of conscious contents from lesions and nonspecific cortical damage, direct brain stimulation and recording, and functional brain-imaging methods that compare conscious vs. unconscious stimulation, like binocular rivalry, which is especially clear in single-cell work in human epileptics (Cerf et al., 2010; Kreiman, Fried, Koch, & 2002; Kreiman, Koch, & Fried, 2000a, 2000b; Reddy, Quiroga, Wilken, Koch, & Fried, 2006). Evidence for brain mechanisms corresponding to unconscious ("id") impulses (e.g., activation of subcortical structures like the amygdala and basal ganglia) and top-down control struggles involving prefrontal regions (e.g., anterior cingulate cortex, dorsolateral prefrontal cortex, orbitofrontal cortex) is very strong. Clinical observation *and* direct evidence strongly support the neural basis for a range of defense mechanisms, so in this respect Freud was on the right track (Berlin, 2011).

There is some variation on how consciousness per se is defined, but consensus from most experts in the field of consciousness research is that consciousness is simply "first-person subjective experience." With this definition in mind, I agree that the hydranencephalic children cited in Solms's article are displaying *some* behaviors that appear to be in direct response to environmental stimuli, but we have no way of knowing whether those behaviors are simply reflexive or whether they are imbued with consciousness. We cannot simply make the assumption that they are conscious. Solms may be arbitrarily labeling unconscious emotions as conscious. Alternatively, due to neuroplasticity as a result of having no cortex *in utero*, the brains of these hydranencephalic children may have reorganized in such a way that some subcortical structures have taken on cortical functions. So what the evidence might show (assuming my distinction between reflex-like behavior and conscious awareness is met) is that consciousness can develop in the absence of much of the forebrain. But it does *not* show that, in a normal brain, consciousness originates anywhere other than in the corticothalamic system.

If Solms's radical theory is correct, it would have an enormous impact on the way we view the brain. We would be forced to assume that people on life support with no cortical activity—that is, brain dead, but with their brainstem intact—are still conscious. If losing awareness and certain cortical functions does not mean losing consciousness, it would necessitate keeping PVS patients alive indefinitely. Since acceptance of Solms's theory as fact would have major practical implications, we must tread lightly and only take on such assumptions as fact once the balance of the evidence is in its favor, which is currently not the case. Solms's article, although provocative, runs afoul of an important scientific dictum: "extraordinary claims require extraordinary evidence" (Sagan, 1980).

REFERENCES

Berlin, H. A. (2011). The neural basis of the dynamic unconscious. *Neuropsychoanalysis,* 13 (1): 5–31.

Boly, M., Faymonville, M. E., Peigneux, P., Lambermont, B., Damas, P., Del Fiore, G., et al. (2004). Auditory processing in severely brain injured patients: Differences between the minimally conscious state and the persistent vegetative state. *Archives of Neurology,* 61 (2): 233–238.

Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M. A., Noirhomme, Q., et al. (2012). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *Journal of Neuroscience,* 32 (20): 7082–7090.

Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R. Q., Koch, C., et al. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature*, 467 (7319): 1104–1108.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375: 121–123.

Esteves, F., Parra, C., Dimberg, U., & Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology,* 31: 375–385.

Fischman, W. M. (1989). Relationship between self-reported drug effects and their reinforcing effects: Studies with stimulant drugs. *NIDA Research Monograph*, 92: 211–230.

Harlow, J. M. (1848). Passage of an iron rod through the head. *Boston Surgical and Medical Journal*, 39: 389–393.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts & Company.

Kreiman, G., Koch, C., & Fried, I. (2000a). Imagery neurons in the human brain. *Nature,* 408 (6810): 357–361.

Kreiman, G., Koch, C., & Fried, I. (2000b). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, 3 (9): 946–953.

Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences USA*, 99 (12): 8378–8383.

Lamb, R. J., Preston, K. L., Schindler, C. W., Meisch, R. A., Davis, F., Katz, J. L., et al. (1991). The reinforcing and subjective effects of morphine in post-addicts: A dose-response study. *Journal of Pharmacology and Experimental Therapeutics,* 259: 1165–1173.

Landsness, E., Bruno, M. A., Noirhomme, Q., Riedner, B., Gosseries, O., Schnakers, C., et al. (2011). Electrophysiological correlates of behavioral changes in vigilance in vegetative state and minimally conscious state. *Brain*, 134 (8): 2222–2232.

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences*, 9 (12): 556–559.

Laureys, S., Faymonville, M. E., Degueldre, C., Fiore, G. D., Damas, P., Lambermont, B., et al. (2000). Auditory processing in the vegetative state. *Brain*, 123 (8): 1589–1601.

Laureys, S., Faymonville, M. E., Peigneux, P., Damas, P., Lambermont, B., Del Fiore, G., et al. (2002). Cortical processing of noxious somatosensory stimuli in the persistent vegetative state. *NeuroImage*, 17 (2): 732–741.

Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., et al. (1999). Impaired effective cortical connectivity in vegetative state: Preliminary investigation using PET. *NeuroImage*, 9 (4): 377–382.

Laureys, S., Lemaire, C., Maquet, P., Phillips, C., & Franck, G. (1999). Cerebral metabolism during vegetative state and after recovery to consciousness. *Journal of Neurology, Neurosurgery, and Psychiatry*, 67 (1): 121.

Laureys, S., Owen, A. M., & Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurology,* 3 (9): 537–546.

Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences.,* 30: 63–134.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., et al. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, 11;316 (5826): 904–906.

Pfaff, D. (2006). *Brain Arousal and Information Theory: Neural and Genetic Mechanisms.* Cambridge, MA: Harvard University Press.

Reddy, L., Quiroga, R. Q., Wilken, P., Koch, C., & Fried, I. (2006). A single-neuron correlate of change detection and change blindness in the human medial temporal lobe. *Current Biology*, 16 (20): 2066–2072.

Sagan, C. (1980). Encyclopaedia Galactica. *Cosmos: A Personal Voyage* [TV series, episode 12]. PBS, 14 December.

Schiff, N. D. (2004). The neurology of impaired consciousness: Challenges for cognitive neuroscience. In: *The Cognitive Neurosciences III*, ed. M. S. Gazzaniga. Cambridge, MA: Bradford Books, pp. 1121–1142.

Schiff, N. D. (2008). Central thalamic contributions to arousal regulation and neurological disorders of consciousness. *Annals of the New York Academy of Sciences,* 1129: 105–118.

Schiff, N. D., Ribary, U., Moreno, D. R., Beattie, B., Kronberg, E., Blasberg, R., et al. (2002). Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. *Brain*, 125 (6): 1210–1234.

Shermer, M. (2011). *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies—How We Construct Beliefs and Reinforce Them as Truths*. New York: Holt.

Shewmon, D., Holmse, D., & Byrne, P. (1999). Consciousness in congenitally decorticate children: Developmental vegetative state as a self-fulfilling prophecy. *Developmental Medicine and Child Neurology,* 41: 364–374.

Velly, L. J., Rey, M. F., Bruder, N. J., Gouvitsos, F. A., Witjas, T., Regis, J. M., et al. (2007). Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia. *Anesthesiology*, 107 (2): 202–212.

Villablanca, J. R. (2004). Counterpointing the functional role of the forebrain and of the brainstem in the control of the sleep-waking system. *Journal of Sleep Research*, 13 (3): 179–208.

Is the Brainstem Really Sufficient for a Consciousness That Would Have Interested Freud?

Commentary by Robin Carhart-Harris (London)

This is an interesting Target Article from Mark Solms, and I am grateful for having been asked to comment on it. However, I have to disagree with some of its main arguments—for example, that the brainstem is sufficient for a form of consciousness that is relevant to the Freudian model of the mind (which depends on the notion of conflict between the ego and the id), and that the id can be associated with the upper brainstem, and the ego with the cortex. I have argued that these ideas seem too crude and are lacking in a sufficient evidence base.

Keywords: brainstem; consciousness; cortex; ego; Freud; id

As I have understood it, Mark Solms's main argument is that the upper brainstem is sufficient for consciousness and that organisms with this and associated limbic structures possess a quality of consciousness associated with the Freudian id but lack the quality of consciousness associated with the Freudian ego. With regards to the latter, Solms proposes that we should look to the cortex—with an emphasis on the somatosensory and motor cortices.

Robin Carhart-Harris: Neuropsychopharmacology Unit, Division of Brain Sciences, Imperial College London, U.K.

The important points put forward in this article are not about how the brain is functionally organized, but how its functional organization maps onto the Freudian model. Another central point is the argument that the structures that best relate to the Freudian id are sufficient to give rise to consciousness, albeit of a primitive sort. Solms also argues that the primary sensory cortices, and related association cortices, generate a representation of the body as an object, whereas upper-brainstem and limbic structures represent the body's internal state on a subjective, interoceptive level. In what follows, I will comment on those aspects of the article I considered to be the most important.

1. Is there a sufficient evidence base for the demarcation lines shown in Figure 1 in the Target Article? For example, to associate the entire prefrontal cortex (PFC) with one system is simplistic. It is well known that the PFC is functionally heterogeneous (e.g., Beckmann, DeLuca, Devlin, & Smith, 2005; Catani et al., 2012; Smith et al., 2009).
2. Freud said many things about the ego, but much of his most highly regarded work focused on its dynamic properties; in particular, its relationship with the id. What is important about the ego is that it is sufficiently distinct from another system in the mind, the id. Granted, the two are connected and related, with the ego having "emerged from the id," but if the two are not sufficiently different, then we would not need to talk about two systems or psychological conflict, we could just talk about one mind—as does cognitive psychology. But Freud dissected the human mind into two systems, a primitive system and a more recent one. The recent system tries to suppress the influence of the older one—but often with the generation of conflict. Freud's comments about the bodily ego are actually few and far between, and when it comes to mapping the ego, we should probably concentrate on what is most important about it—for example, that it refers to our sense of self ("the I") and that it sits atop, and sometimes in conflict with, a more primitive system. With regards to the neurobiological substrates of these systems, I have argued before that we should look to the default mode network (DMN) (Raichle et al., 2001) and the limbic system—and I strongly maintain this view. The DMN contains evolutionarily recent (Van Essen & Dierker, 2007) cortical regions that are strongly connected with the evolutionarily ancient limbic system (Vincent et al., 2006). The DMN is hierarchically superordinate to the limbic system and exerts top-down control on it—while also being driven by it (Carhart-Harris & Friston, 2010).
3. What is the evidence that self-reflection rests on activity in the sensorimotor cortex? If we should look here for the ego, should not self-reflection and other ego-related functions activate these regions? A wealth of evidence suggests that ego functions rest on activity in the DMN (Qin & Northoff, 2011). For example, the DMN consumes more energy (Raichle & Snyder, 2007), receives more perfusion (Zou, Wu, Stein, Zang, & Yang, 2009), and is more widely connected (Hagmann et al., 2008) than other cortical regions. It undergoes significant ontogenetic development (Supekar et al., 2010), it underwent significant evolutionary expansion (Van Essen & Dierker, 2007), and its connectivity has been found to relate to personality (Adelstein et al., 2011) and psychopathology (Whitfield-Gabrieli & Ford, 2012). DMN regions are functionally removed from sensory processing (Sepulcre, Sabuncu, Yeo, Liu, & Johnson, 2012) and are, instead, concerned with high-level functions such as self-consciousness (Qin & Northoff, 2011), complex mental imagery (Hassabis & Maguire, 2009), mental time travel (Buckner & Carroll, 2007), theory-of-mind (Spreng, Mar, & Kim, 2009), and metacognition (Fleming, Weil, Nagy, Dolan, & Rees, 2010). Moreover, reports of "ego-dissolution" are common among users of "psychedelic" drugs such as LSD and psilocybin (magic mushrooms); also, decreased perfusion, venous oxygenation, functional connectivity (Carhart-Harris et al., 2012), and neural synchrony (Muthukumaraswamy & Carhart-Harris, in press) have all been observed in the psilocybin-induced psychedelic state. These findings support the hypothesis that it is from the DMN's self-organized activity that what we refer to as "ego" (i.e., our sense of having a distinct identity or "I") emerges (Carhart-Harris & Friston, 2010; Carhart-Harris, Mayberg, Malizia, & Nutt, 2008).
4. The hydranencephaly cases are interesting, and there seems no reason to doubt that these patients possess a rudimentary form of consciousness, with a varied emotional repertoire. However, it is unclear what cortex is missing and what is present. The images in Figure 3 in the Target Article show some cortical structures present—for example, the inferior temporal cortex. More importantly, however, there are more than just upper-brainstem structures present.
5. Solms's claim that "the state of consciousness as a whole is generated in the upper brainstem" is too strong. There is little question that the brainstem is involved in arousal, but feedback structures and additional drivers are also involved. For example, driving input to the cortex is also provided by the thalamus. This is supported by the finding that thalamic stimulation can be effective for disorders of consciousness (Shah & Schiff, 2010).
6. The argument that the brainstem nuclei provide "the core quality of subjective being" is too strong. Rather, this is likely to be an emergent property of activity in a complex, hierarchically organized system (Deco, Rolls, & Romo, 2009; Eguiluz, Chialvo, Cecchi, Baliki, & Apkarian, 2005; Friston & Ao, 2012; Sporns, 2011; Tononi, 2008). The important

point is that we cannot attribute function to a single structure in isolation; it can only be understood as part of system. Solms probably agrees with this.

7. This may be an aside, but I would question that primary-process cognition is *primarily* about wish-fulfillment, as I would question that dreams are wish-fulfillments. This is one of the weaker points of Freudian theory, in my opinion. If we accept that primary-process cognition is a *primitive mode of cognition*, then it is more defensible to say that it is a mode of cognition that does not effectively reality-test, and that it is *biased* by fears and desires. That is, primary-process cognition does not optimally model the world, because its models are too rough-and-ready; they do not take the time to properly sample the world so to model it faithfully. Sometimes the models are biased by desires, and other times they are biased by anxieties. Freud may have placed too much weight on the former in relation to dreaming. It may be that dreams operate to reinterpret personal reflections in a more favorable, ego-syntonic light—which explains, for example, the dream of Irma's injection (Freud, 1900, p. 107)—but the idea of wish-fulfillment seems to have become more associated with drives—and I think this might be misleading.
8. The treatment of Friston's free-energy theory is nice.

In summary, I agree with the notion that the id is the primary stuff of which our minds are made; however, I would not associate this with the upper-brainstem structures. Instead, it makes more sense to look at the extended limbic *system* and its characteristic neurophysiology (Carhart-Harris, 2007). Nor would I associate the ego with primary- and higher level sensory cortices. Instead, evidence suggests we should look at the relationship between the limbic system and the DMN (Carhart-Harris & Friston, 2010; Laxton et al., 2010), just as Freud looked at the relationship between the ego and the id. I would also question the usefulness of talking about the id as being conscious. Consciousness is a difficult term for which there is no agreed upon definition. One of the most popular definitions focuses on *access*—that is, the ability to report that you are conscious of something (Block, 2009). It may be that the id and the biological processes on which it rests are not sufficient for such metacognition (Fleming, Dolan, & Frith, 2012). However, they may be sufficient for what is referred to as phenomenal consciousness (Block, 2009) or the capacity to have some subjective experience, however rudimentary. Whether we call a subjective experience "conscious" or not may be a matter for the philosophers to debate. Because of disagreements about definitions, this is not an easy issue for science to resolve.

In developing a biologically informed metapsychology, a key thing to consider is that there exist in the mind two *discrete* systems, a primitive one (the id) and a more recent one (the ego). The ego provides some control over the id but is by-and-large "at its mercy." If these systems were not discrete—that is, if they were continuous—then we would not require two terms to describe them. But if we accept that the id and the ego are discrete, then this would imply that the mind and brain underwent an important *phase transition* in its evolution in which it acquired an important new functional capacity in the form of "the ego" and its capacity for metacognition or self-reflection.

Perhaps the best place to start when aiming for a biological account of the mind is with these two sufficiently discrete systems, the id and the ego, characterizing them biologically and highlighting where and how they are different. I would argue that we should focus on the DMN and its relationship with the limbic system. Focusing on the brainstem and cortex may be too general and unspecific to the psychoanalytic model of the mind.

REFERENCES

Adelstein, J. S., Shehzad, Z., Mennes, M., Deyoung, C. G., Zuo, X. N., Kelly, C., et al. (2011). Personality is reflected in the brain's intrinsic functional architecture. *PLoS One,* 6 (11): e27633.

Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B. Biological Science*, 360 (1457): 1001–1013.

Block, N. (2009). Consciousness and cognitive access. *Neuroscience Research*, 65: S28–S28.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences,* 11 (2): 49–57.

Carhart-Harris, R. (2007). Waves of the unconscious: The neurophysiology of *dreamlike* phenomena and its implications for the psychodynamic model of the mind. *Neuropsychoanalysis,* 9 (2): 183–212.

Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., et al. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences USA*, 109 (6): 2138–2143.

Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain,* 133 (4): 1265–1283.

Carhart-Harris, R. L., Mayberg, H. S., Malizia, A. L., & Nutt, D. (2008). Mourning and melancholia revisited: Correspondences between principles of Freudian metapsychology and empirical findings in neuropsychiatry. *Annals of General Psychiatry,* 7 (9).

Catani, M., Dell'acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., et al. (2012). Short frontal lobe connections of the human brain. *Cortex,* 48 (2): 273–291.

Deco, G., Rolls, E. T., & Romo, R. (2009). Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, 88 (1): 1–16.

Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. (2005). Scale-free brain functional networks. *Physical Review Letters*, 94 (1).

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B. Biological Science*, 367 (1594): 1280–1286.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science,* 329 (5998): 1541–1543.

Freud, S. (1900). *The Interpretation of Dreams*. *Standard Edition*, 4/5.

Friston, K., & Ao, P. (2012). Free energy, value, and attractors. *Computational and Mathematical Methods in Medicine*. doi:10.1155/2012/93786

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology,* 6 (7): e159.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B. Biological Science*, 364 (1521): 1263–1271.

Laxton, A. W., Tang-Wai, D. F., McAndrews, M. P., Zumsteg, D., Wennberg, R., Keren, R., et al. (2010). A phase I trial of deep brain stimulation of memory circuits in Alzheimer's disease. *Annals of Neurology,* 68 (4): 521–534.

Muthukumaraswamy, S., & Carhart-Harris, R. (in press). Cortical desynchronisation underlies the human psychedelic state. *Nature Neuroscience*.

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, 57 (3): 1221–1233.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences USA*, 98 (2): 676–682.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, 37 (4): 1083–1090 [discussion 1097–1089].

Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *Journal of Neuroscience,* 32 (31): 10649–10661.

Shah, S. A., & Schiff, N. D. (2010). Central thalamic deep brain stimulation for cognitive neuromodulation—a review of proposed mechanisms and investigational studies. *European Journal of Neuroscience*, 32 (7): 1135–1144.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences USA*, 106 (31): 13040–13045.

Sporns, O. (2011). The human connectome: A complex network. *Annals of the New York Academy of Sciences*, 1224: 109–125.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, 21 (3): 489–510.

Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, 52 (1): 290–301.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215 (3): 216–242.

Van Essen, D. C., & Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron,* 56 (2): 209–225.

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., et al. (2006). Coherent spontaneous activity identifies a hippocampal–parietal memory network. *Journal of Neurophysiology*, 96 (6): 3517–3531.

Whitfield-Gabrieli, S., & Ford, J. M. (2012). Default mode network activity and connectivity in psychopathology. *Annual Review of Clinical Psychology*, 8: 49–76.

Zou, Q., Wu, C. W., Stein, E. A., Zang, Y., & Yang, Y. (2009). Static and dynamic characteristics of cerebral blood flow during the resting state. *NeuroImage*, 48 (3): 515–524.

Beyond the Reward Principle: Consciousness as Precision Seeking

Commentary by Aikaterini Fotopoulou (London)

I use an influential computational theory of brain function—the free-energy principle—to suggest three points of added complexity to Mark Solms's intriguing descriptions of the embodied mind: (1) The link between ego and cognitive automaticity is not as straightforward as Solms suggests; instead, cognition strives for both inference and *flexibility* in relation to the changing world and the inflexible drives. (2) Affective consciousness may primarily map the degree of *uncertainty* (not pleasure) of internal bodily signals; subcortical areas are the neurobiological source of this facet of consciousness that in itself is likely to be localized among many, distributed brain areas. (3) Our innate motivational systems—the id—ultimately serve the same optimization principle as the ego; however, unlike the latter, they call for automaticity in behavior, on the basis of innate unconscious priors that are fulfilled by instinctual "e-motions" and other reflexes, understood as evolutionarily defined, primitive forms of active and perceptual inference.

Keywords: consciousness; embodiment; emotion; free energy; motivation; predictive coding

Mark Solms's rich and provocative article weaves together classic concepts of Freudian metapsychology and insights from affective neuroscience into a novel, lucid neuropsychoanalytic account of embodiment and consciousness. Doing justice to the many research traditions and creative links that the article invokes is not possible in this brief commentary. Moreover, I wish to take nothing away from the force and clarity by which Solms contrasts the subjectively felt body to the "cognitivized" one and questions the simplistic equation of consciousness with ego. In this commentary, I will use a computational theory of brain function—the free-energy framework (Friston, 2005)—to merely suggest three points of added complexity to Solms's intriguing descriptions of the embodied mind. These will relate to (1) the nature of the ego (our cognition); (2) the distinction between phenomenal and perceptual consciousness; and (3) the id (our innate drives).

A theoretical framework from computational neuroscience

The starting point of the free-energy framework (Friston, 2005) is that the world is an uncertain place for self-organizing biological agents to survive. This inherent ambiguity of the world threatens our need to occupy a limited repertoire of sensory states (e.g., humans need certain ranges in the environmental temperature in order to survive). If however we cannot predict the causes of possible changes in the world with any certainty, we may find ourselves in surprising states for longer periods than those we could biologically sustain. We thus come up with a defiant solution. We base our predictions about our sensory states on unconscious inferences about their causes in the world (von Helmholtz, 1866). On the basis of limited or noisy information, our brain engages in probabilistic representations of the causes of our future states in an uncertain world so that it maintains hypotheses ("generative models") of the hidden causes of sensory input. Theoretical neuroscientists use Bayesian theory to formalize this kind of inference and a number of other computational terms about probability distributions, such as "free energy," "uncertainty" and "surprise," that have formal (mathematical) definitions. Here I attempt to find faithful "psychological translations" for some of these concepts in order to examine the ideas put forward by Solms within this "psychologized" version of the free-energy framework.

According to the framework, the brain attempts to reduce the probability of being surprised by the world by reducing its own representational errors over time. These errors have been conceptualized as free energy, on the basis of the formal definition of the latter: a quantity from informational theory that bounds (is greater than) the evidence for a model of data (Feynman, 1972; Hinton & van Camp, 1993). When the data is sensory, free energy bounds the negative log-evidence (surprise) inherent in them, given a model of how the data were caused. Furthermore, our brain is assumed to achieve the minimization of free energy by recurrent message passing among hierarchical levels of cortical systems, so that various neural subsystems at different hierarchical levels minimize uncertainty about incoming information by structurally or functionally embodying a prediction (or a prior) and responding to errors (mismatches) in the accuracy of the prediction, or prediction errors (Rao & Ballard, 1999). Such message passing is considered neurobiologically plausible on the basis of functional asymmetries in

Aikaterini Fotopoulou: King's College London, Institute of Psychiatry, London, U.K.

cortical hierarchies (e.g., Mesulam, 2012). Minimizing free energy corresponds to explaining away prediction errors following the principles of Bayes (Friston, 2010).

However, perceptual inference cannot take us far in terms of our ultimate goal: surviving in an uncertain world. Psychologically speaking, we may become better in predicting ("mentalizing") the changes in the environment that act to produce sensory impressions on us, but we cannot on this basis change the sensations themselves and hence ultimately their surprise. It is only by acting upon the world that we can "resample" the world to ensure that we satisfy our predictions about the sensory input we expect to receive. Thus, action is understood as being elicited to fulfill prior expectations about proprioceptive sensations, not desired sensory states (as optimal motor-control theory suggests). Both action and perception are governed by the imperative of free-energy minimization: action reduces free energy by changing sensory input, while perception reduces free energy by changing predictions.

At this point, we can follow Solms (see also Carhart-Harris & Friston, 2010; Fotopoulou, 2012; Hopkins, 2012) in considering this framework in parallel with the central Freudian "economic" concepts of "free" and "bound" energy. Although the Freudian and Fristonian notions of free energy differ between them, they were both inspired by Helmholtzian ideas about thermodynamics and commonly convey the imperative to "bind" and ultimately "minimize" a quantity (formal surprise in Friston; nervous excitation and then psychic energy in Freud) that otherwise renders the functioning of a biological system suboptimal in an unpredictable world. As described above, the free-energy framework places emphasis on an antagonism between an unpredictable world and a predicting mind. Freud, however, placed emphasis on an antagonism between *an inflexible body* and an unpredictable world; he claimed that the mind (the ego) is formed on the basis of an antagonism between the organism's biological needs (and corresponding inherited drives, the id) and an unaccommodating world. Of course, it should be clear that the two frameworks bear on the same single, ultimate principle; a biological system with given constraints needs to reduce its modes of exposure to the unpredictable world in order to maintain and prolong its evolutionarily constrained existence (e.g., Friston, 2011; Friston & Ao, 2012). I fear, however, that Solms's new division of labor between the ego and the id, and his denigration of drives (innate motivational priors) to affective consciousness and of ego (cognition) to mere representation and automatization, risks de-emphasizing the central place of the Freudian antagonism between an inflexible body and an unpredictable world. In fact, Solms says that much when he regards his new conscious, subcortical id as the seat of the novel, the salient, and the emotional in the brain and his unconscious, cortical ego as the driver of automaticity.

The flexible ego and consciousness as precision seeking

In the free-energy framework, the challenge of the organism is to navigate the world by sustaining a set of prior beliefs, sufficiently robust that the organism does not react reflexively to incoming sensory stimuli. At the same time, and contrary to what Solms claims about automaticity (Section 7), our generative models of the world must not be so immutable that our responses become fixed, stereotypical, and insensitive to unpredictable change. Indeed, an intrinsic component of the free-energy framework is that our generative models need to maintain an optimal, dynamic balance between their robustness and flexibility. In Bayesian terms, organisms need to probabilistically infer two properties of the world: its states (*content*; mathematically this can be thought of as the center of a probability distribution) and the uncertainty (*context*; the dispersion of such distribution) about such states. It is perhaps Solms's apparent disregard of the latter that leads him to equate the ego with the driver of automaticity and to claim that the reduction of salience constitutes one of the aims of the ego (Section 7). Increases in salience, novelty, and motivational value do not oppose the principle of minimization of free energy. In fact, the opposite applies: optimal inference in both perception and action requires optimizing the precision (mathematically inverse dispersion or variance, and hence the inverse of uncertainty) of sensory signals (Feldman & Friston, 2010; Friston, Shiner, et al., 2012). Uncertainty is thought of as encoded mainly by synaptic gain that encodes the precision of random fluctuations about predicted states. It follows that neuromodulations of synaptic gain (such as dopamine and acetylcholine) do not signal (reward or pleasure) prediction errors about sensory data but the context in which such data were encountered. In other words, such neuromodulators report the salience of sensorimotor representations encoded by the activity of the synapses they modulate. This is important, especially in hierarchical schemes, where precision controls the relative influence of bottom-up prediction errors and top-down predictions.

In psychological terms, the processing of salience expectancy allows the organism to control the significance it attributes to the sensory data it uses to update its predictions or to explain away prediction errors. As regards exteroception, this processing of salience can be seen as attention in perceptual inference (Feldman & Friston, 2010), and as affordance (latent action possibilities of cues in the environment) in active inference (Friston, Shiner, et al., 2012). In interoception, optimizing the precision of internal body signals can be seen as increased *interoceptive sensitivity and related feelings of arousal* in perceptual inference (note, however, that this is not synonymous with increased prediction error about interoceptive signals; see below) and as increased *seeking behaviors* in active inference (see also Friston, Adams, Perrinet, & Breakspear, 2012). Understanding the "objectless," so-called SEEKING system (Panksepp, 1998) as the driver of a kind of enacted search for increased precision regarding internal body priors fits with what we know about the neurobiology of dopamine and related, bottom-up neuromodulators (Friston, Shiner, et al., 2012; Pfaff & Fisher, 2012). Viewing the SEEKING system as supporting precision seeking also has intuitive meaning: we are motivated to sample the world when we do not know where surprise will come from, and vice versa (Anselme, 2010).

One core aspect of consciousness may serve to register the aforementioned quality of "uncertainty" and its inverse quality, precision. This view goes against the intuitive, long-standing view of core affective consciousness as monitoring hedonic quality, expressed by Solms in Freudian terms as the pleasure–unpleasure series. Instead, I propose that the core quality of this aspect of consciousness (as opposed to perceptual consciousness; see below) is a kind of certainty–uncertainty, or disambiguation principle. Certainty in this sense is not synonymous with prediction—that is, it is not a measure of what was predicted, nor what occurred. Nor is it first and foremost the mental process that tells us what is good or bad for us homeostatically (although because of our innate constraints this is one common derivative of the certainty–uncertainty principle; see discussion of drives below). In this sense, consciousness is the process that tells us that we feel increased desire for and show approach tendencies toward unfamiliar, exotic, and unpredictable foods, destinations, and sexual partners not because we are predicting particularly rewarding experiences, but, rather, because we cannot predict such experiences with sufficient certainty. As I will argue, it is instinctual "e-motions" (inherited psychomotor patterns) and other innate priors (the Freudian id) that oppose this uncertainty principle and, instead, call for a relative automaticity and inflexibility in the system. If we leave ego to its own devices, including its capacity for both perceptual inference and conscious disambiguation, it will lead the organism not to automaticity but, rather, to a never-ending and ultimately resource-draining, self-destructive cycle of seeking and cognitively finding (predicting and learning) of endless random fluctuations in the environment.

Before returning to the unconscious id, however, it is worth mentioning that a second type of consciousness can be conceived. Perceptual consciousness, both interoceptive and exteroceptive, may be instantiated as an instance of otherwise unconscious processes of perceptual inference about the causes of sensations. Indeed, it has recently been proposed that subjective feeling states arise from predictive inferences on the causes of interoceptive signals (Seth, Suzuki, & Critchley, 2011). This "interoceptive predictive coding" model is compatible with the so-called James–Lange theory of emotions (James, 1890; Lange, 1885) to the degree that it claims that feelings are understood to arise from *perceptions* of physiological changes. Starting with the precise interpretation of James's work, classic debates in psychology have unfolded about whether bottom-up, direct bodily signals and/or top-down cognitive representations, categories, or evaluations of physiological changes are responsible for feeling states. This model can specify the dynamic balance between bottom-up and top-down signals in interoception at various hierarchical levels, yet the interoceptive bodily self in this theory is always an inference (like Solms's objective body)—that is, it is inferred on the basis of generative models about the likely causes of one's interoceptive signals.

Contrary to Seth, Suzuki, and Critchley, Solms views the core of affective consciousness as nonrepresentational. As I proposed above, this aspect of consciousness can be best characterized as interoceptive sensitivity and precision seeking. Moreover, the dynamic source of affective consciousness and its most raw psychological manifestations may well depend on activity in the upper brainstem and limbic areas that Solms mentions. This implies a certain degree of functional segregation or modularity (for discussion, see Fotopoulou, in press) and, indeed, as Solms's suggests, a given hierarchy between more raw aspects of affective consciousness and more cognitivized aspects of consciousness. Nevertheless, the neural basis of the various affective qualities of consciousness is most likely generated at multiple and different levels of the hierarchy due to the functional integration (Friston, 1994) or the synchronization (Engel, Fries, & Singer,

2001) of *activity between such areas and cortical areas.*

The inflexible drives and instinctual emotions as primitive forms of active inference

This section stresses that our inherited motivational systems should not be equated with affective consciousness or salience and, moreover, that it is the drives (the Freudian id, not the ego) that call for a relative automaticity in both cognition and behavior. Interestingly, this was the very point that Freud put forward about drives and the Nirvana principle in 1920. The aim of minimizing free energy (and hence surprise) is to ensure that agents spend most of their time in a small number of "valuable" states. Valuable states are not first and foremost conscious, pleasurable states, as Solms implies (and Freud thought until 1920), but unsurprising states—that is, states that evolution informs us our species most frequently occupied. Value, like free energy, depends on an organism's generative model and its implicit, heritable priors, optimized at different, evolutionary time scales; their job is to specify the innate value of certain attractive sensory states. These expectations thus include the prior that the organism itself (as part of the environment) occupies an invariant (attracting) set of physical (including internal) states. Valued states are therefore expected states. In other terms, evolution equips an organism with optimized prior expectations about the states the organism is likely to encounter (these ideas are related to neural Darwinism; Friston, 2010).

However, as mentioned above, as priors are mere hypotheses, the agent is evolutionarily primed to test them by using sensory samples from the environment. Our primary expected states are therefore specified genetically, but in one's lifetime they are fulfilled behaviorally, under active inference. Unlike the more object-less, exploratory SEEKING system mentioned above (Panksepp, 1998), the other instinctual, object-specific, primary e-motions described by Panksepp (1998) seem to fit exactly the role of primitive active inference in relation to innate priors. Reflexive, sensorimotor patterns are elicited to fulfill prior expectations about attractive sensory states of the organism, in the same way that classic reflexes elicit movement to fulfill prior expectations about proprioceptive sensations. In Freudian terms, it is the id that calls for a relative automaticity and reduction of states to a minimum. This minimization of nonevolutionarily subscribed sensory states seems to be the ultimate guiding principle of our drives (innate priors), rather than the pleasure principle (homeostatically rewarding values). Indeed, as Freud suggested in 1920, the pleasure principle seems to be secondary to this minimization imperative that governs the id (the Nirvana principle that Solms now attributes to the ego). In Friston's words, "the problem of finding sparse rewards in the environment is nature's *solution* to the problem of how to minimize the entropy (average surprise or free energy) of an agent's states: by ensuring they occupy a small set of attracting (that is, rewarding) states" (Friston, 2010, p. 135; emphasis added). It thus falls upon the ego—or cognition—to tailor this inflexible, inherited minimization imperative to the demands of the unpredictable world during one's lifetime. Under perceptual and active inference, the ego thus builds empirical priors on the foundations of innate priors. The ego's "cognitivized" generative models allow for a more flexible and efficient, yet motivationally constrained relation with the ambiguous world. This includes retaining an optimal degree of instability in perceptual inference that allows the ego to explore alternative hypotheses about the causes of sensory states (Friston, Breakspear, & Deco, 2012). Thus, while the world is ambiguous and potentially surprising and the id strives to minimize the states that the organism encounters to the very few that would satisfy basic, homeostatic needs, the ego strives for an optimal balance between the two.

Conclusion

In summary, I have used an influential computational theory of brain function—the free-energy principle—to suggest three points of added complexity to Solms's intriguing descriptions of the embodied mind: (1) Most of the ego may well be unconscious, but the link between ego and cognitive automaticity is not as straightforward as Solms suggests. Instead, cognition strives for both inference *and flexibility* in relation to the changing world and the inflexible drives. (2) Affective consciousness may primarily map the degree of *uncertainty* (not pleasure) of internal bodily signals. Subcortical areas are the neurobiological *sources* of this facet of consciousness that in itself is likely to be localized between many, distributed brain areas. Finally, (3) our innate motivational systems—the id—ultimately serve the same optimization principle as the ego, but, unlike the later, they call for automaticity in behavior, on the basis of innate unconscious priors that are fulfilled by instinctual e-motions and other reflexes, understood as evolutionary defined, primitive, and inflexible forms of active and perceptual inference.

Of course, the above speculative view of the motivated and embodied mind leaves unanswered more theoretical and empirical questions than those it attempts to answer. At least one important point of complexity I did not touch upon is the role of other agents in both perceptual and "precision-seeking" consciousness. Similarly, I cannot possibly do justice in this brief commentary to complex notions such as "repression" and the "dynamic unconscious." Nevertheless, a few implications could be highlighted. It is easy to infer from what I have written how conflict between the demands of different innate priors (for further discussion, see Hopkins, 2012), as well as between the unconscious id (which seeks to reduce all non-"prescribed" evolutionary states) and the conscious ego (which seeks to represent and learn all novel signals in the internal and external environment) is therefore unavoidable. The conflict between the ego and the id, for example, may be why risk and danger both attract and scare us. It is also easy to see why Freud insisted on an antithesis between unconscious drives and the conscious feelings that originate in relation to them. Drives themselves (innate priors) are unconscious and minimally reflective (they are reflexively fulfilled by instinctual e-motions), and hence they can never be fully "updated" by the ego according to the changes in the external world (perceptual inference and learning). On the contrary, it is important that the ego registers the core feelings that relate to the specificity of such innate predictions (the bottom-up modulation of the certainty of such predictions) so that the cognitive resources available for scanning the world and the body for novelty and salience are always constrained by, and in competition with, the high precision of our innate expectations. These speculative ideas do, of course, require further specification, proper modeling and empirical testing, but I hope they at least hold the potential of contributing some added "precision" to Solms's rich, wide-ranging, and thought-provoking view of the embodied mind.

REFERENCES

Anselme, P. (2010). The uncertainty processing theory of motivation. *Behavioural Brain Research*, 208: 291–310.

Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain*, 133: 1265–1283.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature*, 2: 704–716.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 2 (4): 215.

Feynman, R. P. (1972). *Statistical Mechanics*. Reading: Benjamin.

Fotopoulou, A. (2012). Towards psychodynamic neuroscience. In: *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*, ed. A. Fotopoulou, M. Conway, & D. Pfaff. Oxford: Oxford University Press, pp. 25–46.

Fotopoulou, A. (in press).Time to get rid of the "Modular" in neuropsychology: A unified theory of anosognosia as aberrant predictive coding. *Journal of Neuropsychology*.

Freud, S. (1920). *Beyond the Pleasure Principle. Standard Edition,* 18.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2: 56–78.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences*, 360: 815–836.

Friston, K. (2010). The free-energy principle: A unified brain theory. *Nature Reviews. Neuroscience*, 11: 127–138.

Friston, K. (2011). Policies and priors. In: *Computational Neuroscience of Drug Addiction*, ed. B. Gutkin & S. H. Ahmed. New York: Springer, pp. 237–283.

Friston, K., & Ao, P. (2012). Free-energy, value and attractors. *Computational and Mathematical Methods in Medicine*. doi:10.1155/2012/937860

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3 (151): 1–20.

Friston, K. J., Breakspear, M., & Deco, G. (2012). Perception and self-organized instability. *Frontiers in Computational Neuroscience,* 6 (44): 1–19

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, 8 (1), e1002327.

Hinton, G. E., & van Camp, D. (1993). Keeping neural networks simple by minimising the description length of weights. In: *COLT '93 Proceedings of the Sixth Annual Conference on Computational Learning Theory*. New York: Association of Computing Machinery, pp. 5–13.

Hopkins, J. (2012). Psychoanalysis, representation and neuroscience: The Freudian unconscious and the Bayesian brain. In: *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*, ed. A. Fotopoulou, M. Conway, & D. Pfaff. Oxford: Oxford University Press, pp. 230–265.

Mesulam, M. (2012). The evolving landscape of human cortical connectivity: Facts and inferences. *NeuroImage*, 62 (4): 2182–2189.

Panksepp, J. (1998). *Affective Neuroscience*. New York: Oxford University Press.

Pfaff, D. W., & Fisher, H. E. (2012). Generalized brain arousal mechanisms and other biological, environmental and psychological mechanisms that contribute to libido. In: *From*

the Couch to the Lab: Trends in Psychodynamic Neuroscience, ed. A. Fotopoulou, M. Conway, & D. Pfaff. Oxford: Oxford University Press, pp. 64–84.
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience,* 2: 79–87.
Seth, A. K., Suzuki, K., & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2 (395). doi: 10.3389/fpsyg.2011.00395
von Helmholtz, H. (1866). Concerning the perceptions in general. In: *Treatise on Physiological Optics, Vol. 3* (3rd edition), trans. J. P. C. Southall. New York: Dover, 1962.

Consciousness and Hierarchical Inference

Commentary by Karl Friston (London)

I greatly enjoyed reading Mark Solms's piece on the conscious id. Mindful of writing this commentary, I noted (in the margins of the target article) points of contact between his formulation and our more formal—if somewhat dryer—treatment using Helmholtzian notions of free energy. It was clear, after a few pages, that I was not going to be able to cover every aspect of the remarkable consilience between the two approaches. Instead, I focus on substantiating Solms's key conclusions from the perspective of hierarchical inference in the Bayesian brain.

Keywords: consciousness; free energy; hierarchy; inference; neuronal activity; perception

It strikes me that the neuropsychoanalysis movement—more than any other field—confronts the theoretical challenges that attend affect, emotion, and interoception. While there is an enormous amount of theoretical work on Bayes-optimal perception and motor control in the exteroceptive and proprioceptive domains (Knill & Pouget, 2004; Körding & Wolpert, 2004), there is a curious absence of formal theory pertaining to emotion and interoception. At first glance, one might consider value-learning and optimal decision theory as good candidates for a theory of emotion and affect. However, these normative approaches are rather shallow—appealing tautologically to behaviorist or economic notions such as reward and utility. In what follows, I provide a brief overview of the free-energy principle discussed in the Target Article. This principle provides a framework to revisit the issues of hierarchical representation and conscious and unconscious inference and their location within the cortico–subcortical hierarchy. After considering the neurobiological substrates of conscious inference, I comment briefly on Solms's conclusions about therapeutic interventions.

Karl Friston: The Wellcome Trust Centre for Neuroimaging, University College London, London, U.K.
Acknowledgements: This work was funded by the Wellcome Trust.

Free energy and neurobiology

The free-energy formulation referred to by Solms is an attempt to apply information theory to self-organizing systems like the brain (Friston, 2010). Its premise is simple: to maintain a homoeostatic and enduring exchange with the world (Ashby, 1947), we have to counter perturbations to the states that we expect to be in. In short, we have to minimize surprising violations of our predictions. Mathematically, this surprise cannot be measured directly; however the brain can compute something called *free energy*, which provides a proxy for surprise. Roughly speaking, free energy is prediction error—namely, the mismatch between bottom-up sensations and top-down predictions. These predictions rest on a model of our world that generates predictions in the *exteroceptive*, *proprioceptive*, and *interoceptive* domains. The minimization of exteroceptive prediction error can be cast as perceptual synthesis or inference; the minimization of proprioceptive prediction error corresponds to behavior (as implemented by classical motor reflexes); and the minimization of interoceptive prediction error corresponds to autonomic or visceral homeostasis (mediated by autonomic reflexes). The neuronal substrate of this minimization is probably simpler than one would imagine: a substantial amount of physiological and anatomical evidence suggests that the brain encodes predictions and prediction errors with the neuronal activity of separable neuronal populations. These two populations pass messages to each other, where populations encoding prediction errors (denoted by ξ in Figure 1) drive populations encoding predictions (denoted by μ in Figure 1). In turn, these predicting populations suppress or inhibit prediction-error populations. Crucially, neuroanatomical evidence suggests that the (generative) model used by the brain is hierarchical, such that top-down predictions try to explain or suppress prediction errors in the level below, while bottom-up prediction errors subvert themselves by informing and optimizing the predictions in the level above (Mumford, 1992; Rao & Ballard, 1999). In this setting, hierarchical predictions come to represent the causes of sensory input in a Bayesian sense. This recurrent and hierarchically deployed process reduces prediction errors at all levels of the hierarchy—thereby optimizing a hierarchical representation (dynamic prediction) of the sensorium, with multiple levels of description.

The analogy between this Helmholtzian suppression of free energy and Freudian free energy is self-evident: the binding of free energy (prediction errors) corresponds to a top-down suppression, which necessarily entails an explanation or resolution of violated predictions. Crucially, the hierarchical structure of generative models—and implicit emergence of nervous energy (activity of prediction error populations)—speaks exactly to Freud's deepest insight, which, states Solms, rests upon the "*depth* (or hierarchy) in the mind."

Hierarchical inference

Clearly, lower levels of hierarchical inference are closer to the sensorium and represent more elemental (and transient) causes of sensory input. Conversely, higher levels of the hierarchy can "see" multiple input modalities. At this point, we start to see the structural basis of Solms's dichotomy between the *autonomic* body (representations or predictions of interoceptive input) and the *somatomotor* body (representations of exteroceptive and proprioceptive input), where these domains converge at higher levels (see Figure 1). Put another way, high-level intransigent representations (mental solids) have an amodal aspect and provide bilateral top-down interoceptive and exteroceptive predictions. In this sense, high-level representations have, necessarily, interoceptive attributes. This resonates with the notion that high-level (e.g., executive or second-order) representations are supported by—or derived from—interoceptive representations. It also suggests that affect is an intrinsic property of the brain: Solms states that "Affect may accordingly be described as an interoceptive sensory modality—but that is not all it is."

In terms of hierarchical inference, affect is a construct or attribute of a higher level representation that is used to explain interoceptive inputs at a lower level—in the same sense that color is used to explain wavelength-selective responses in early visual cortex (Zeki & Shipp, 1988). However, the parallel between hierarchical inference and the dichotomy developed by Solms rests upon a mapping between inference and consciousness.

Free energy and consciousness

The original writings of Helmholtz (1866) focused on unconscious inference in the visual domain. However, in hierarchical (deep) inference schemes (Dayan, Hinton, & Neal, 1995), it is tempting to associate probabilistic representations—encoded by the activity of populations encoding predictions—with consciousness. Many of the attributes of consciousness are shared with these probabilistic representations. In brief, these probability distributions (known as *posterior beliefs*) are encoded by their *sufficient statistics*, such as their mean and variance. For example, the posterior mean or *expectation* is encoded by the activity of populations encoding predictions. This is important because it means that a probabilistic representation is induced by biophysical states of the brain—and uniquely associates one representation (consciousness) with one brain. However, the representation is not the biophysical state that induces it—in the sense that a probability distribution is not the same as its mean and variance. Intuitively, this means that I cannot possess your beliefs (consciousness), but I can believe you believe (I can have beliefs about your biophysical states). If one admits a mapping between consciousness and the probability distribution induced by expectations or predictions, then the hierarchical architecture of our brains has profound implications for consciousness and the arguments pursued by Solms.

Where is the top (center) of the hierarchy?

A tenet of Solms's argument is his deconstruction of the corticocentric view of consciousness. He argues (with compelling empirical evidence) that consciousness resides in (or is generated from) upper-brainstem structures, which may be embellished by (or support) cortical elaborations. In what sense is this consistent with hierarchical inference in the brain? Hierarchical

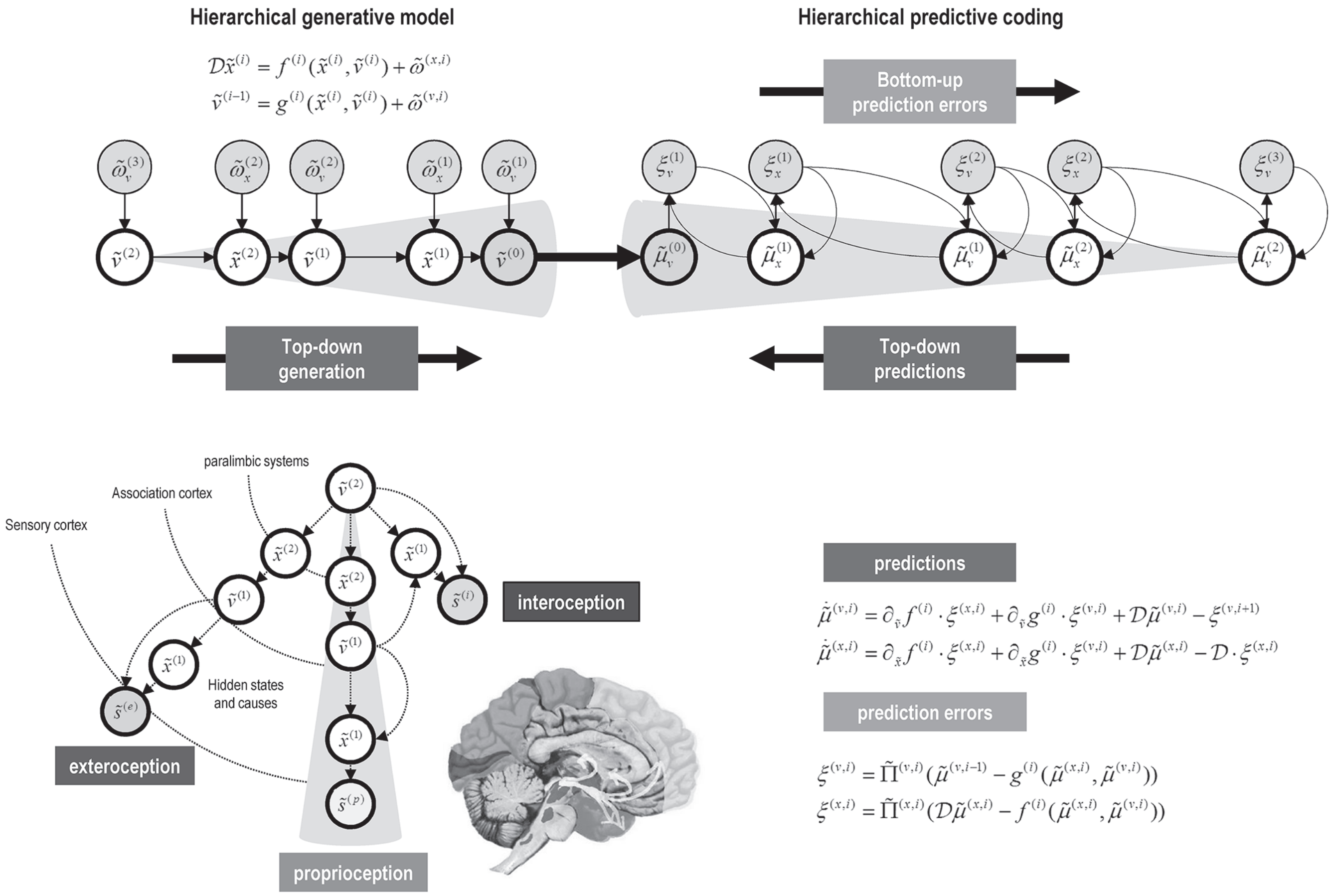

Figure 1. The putative neuronal architectures that might optimize posterior beliefs about the state of the world, using hierarchical generative models. Upper panel: part of a generative model is shown on the left, in terms of a cascade of hidden states and causes in the world that produce sensory input. This architecture is mirrored by hierarchies in the brain that try to explain the input and (implicitly) come to represent the hidden states. Although the details are not important, hidden states $\tilde{x}^{(i)}$ model dynamic dependencies in the way that sensory information is generated, while hidden causes $\tilde{v}^{(i)}$ link hierarchical levels and provide the generative model with a deep structure. The stochastic differential equations (upper left) provide a mathematical specification of the model. The brain infers the hidden causes of sensory input, in terms of posterior expectations or predictions about the hidden states and causes at each level of the hierarchy. The most popular scheme for this hierarchal inference is known as predictive coding, illustrated in the upper right. In this scheme, hidden causes and states are represented in terms of predictions $(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)})$ that are driven by prediction errors $(\xi_x^{(i)}, \xi_v^{(i)})$. Crucially, prediction errors are passed forward to provide bottom-up guidance to neuronal populations encoding predictions, while top-down predictions are assembled to form prediction errors. This process continues until prediction error has been minimized and the predictions become optimal in a Bayesian sense. Lower left panel: this illustrates a model generating exteroceptive, proprioceptive, and interoceptive sensations. Again, the details of this model are not important—it is just meant to illustrate that hierarchical models can have a form in which there is no top but, rather, a center. In this schematic, the scale of grey corresponds to the insert (medial view of the brain): medium grey denotes primary sensory (exteroceptive) input; light grey denotes proprioceptive input; and dark grey denotes interoceptive input. This scheme is based on Figure 1 in the Target Article. The key thing to take from this schematic is that interoceptive parts of the hierarchy are more intimately associated with the center, relative to the peripheral (primary sensory) cortex. Lower right panel: these are the equations that minimize prediction error or free energy (using a gradient descent). They are provided to indicate that the free-energy principle prescribes specific and biologically plausible neuronal dynamics. It can be seen that these equations are based on quantities specified by the generative model and have a relatively simple mathematical form. Of particular note is that the prediction errors are scaled by precision—denoted by $\tilde{\Pi}^{(i)}$. For details about mathematical form and notation, see Friston (2008).

probabilistic representations exist at all levels of the hierarchy. In this context, Solms's arguments make perfect sense, in that different representational attributes can be associated with different locations within the hierarchy. For example, representations with an affective aspect (the id) could be located in systems making interoceptive predictions and coexist (necessarily) with somatomotor representations in the cortex. However, this does not address the question of which "is intrinsically conscious." Let us assume that *intrinsically conscious* means hierarchically supraordinate, in the sense that intrinsic predictions entail extrinsic (exteroceptive and proprioceptive) predictions. So what is the evidence that brainstem regions and associated (para-)limbic brain systems are hierarchically supraordinate to cortical systems? There are two simple lines of evidence—one obvious and one not. It is obvious that primary sensory cortex is at the lowest level of the hierarchy. In other words, there are representations at the cortical level that are only a few synapses away from the sensorium. Clearly, higher order representations of "things and words" that have a temporal persistence involve association cortex. However, to treat the cortex as a functionally homogeneous epicenter is untenable—indeed, it is often depicted on the periphery of centrifugal hierarchies (see Figure 1; see also Mesulam, 1998). The second—less obvious—reason appeals to the inference framework above. In hierarchical inference, top-down predictions fulfill the role of something called *empirical priors* (predictions about predictions). However, at the top (or center) of the hierarchy there are no top-down predictions, and expectations become *full priors*. These expectations are usually associated with the instincts and prior beliefs about bodily states that are selected by evolution (necessary for survival). Neuroanatomically, instinctual or innate priors are concerned with interoceptive inputs and may be entailed by the circuitry and physiology of the upper-brainstem, limbic, and paralimbic systems. This is entirely consistent with the intrinsic representations ascribed to these areas. As discussed in the next section, there is one further reason why these particular systems have an important role in specifying prior (instinctual) beliefs, which touches on the implications for therapy.

Precision, uncertainty, and therapeutic efficacy

Crucially, top-down predictions are not just about the content of lower level representations but also predict their reliability or *precision* (denoted by $\tilde{\Pi}$ in Figure 1). Mathematically, precision is inverse variance or uncertainty. This sort of top-down prediction is thought to be mediated by neuromodulatory mechanisms that optimize the (attentional) gain of populations encoding prediction errors (Feldman & Friston, 2010). This is sensible, in that boosting precise prediction errors gives them a preferential or selective influence on higher (deeper) hierarchical inference. The key thing here is that the precision has itself to be predicted. This means that particular brain systems broadcast posterior beliefs about the precision of various interoceptive and proprioceptive representations—and can, effectively, choose what to explain.

The Bayes-optimal encoding of precision in the brain has already been discussed in terms of attention and affordance and even as an explanation for the emergence of hysterical symptoms (Edwards, Adams, Brown, Pareés, & Friston, 2012). Furthermore, it may provide an interesting metaphor for the repression of (Freudian) free energy, through the neuromodulatory suppression of prediction error units encoding (Helmholtzian) free energy. In the present context, predictions about where precision should be deployed within a hierarchy may be encoded by the activity of classical neuromodulatory transmitter systems that ascend from the extended reticular activating system and upper brainstem—identified in the Target Article. Indeed, at a most basic level, it is this system that controls (through neuromodulatory efferents) the basic cycles of conscious level associated with sleeping and waking (Hobson, 2009). Another classical example is dopamine, which has not only been implicated in neuropsychiatric disorders such as schizophrenia and Parkinsonism but plays a central role in theories of value-dependent learning and emotional behavior (Schultz, 1998). In short, the interior of the brain houses not only the systems necessary for consciousness but also elaborates some of the most important top-down predictions that set the tone for inference elsewhere in the brain—namely, predictions about the precision or salience of prediction errors in one modality (or level) in relation to another.

Therapeutically, as intimated by Solms, locating an intrinsically conscious (representational) capacity at the subcortical and paralimbic level may have important implications for therapy. The nice thing here is that viewing the brain as an inference machine (Dayan, Hinton, & Neal, 1995) means that one can easily motivate therapeutic interactions in terms of changing (posterior) beliefs. At the same time, one can understand this optimization in the context of how we represent precision or uncertainty and the role of key neurotransmitter systems such as the dopaminergic system. In one sense, the therapeutic relationship may provide, as Solms states, the (unattainable) state of Nirvana "that we now learn, to our surprise, is what the ego aspires to."

REFERENCES

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *Journal of General Psychology*, 37: 125–128.

Dayan, P., Hinton, G. E., & Neal, R. (1995). The Helmholtz machine. *Neural Computation*, 7: 889–904.

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. J. (2012). A Bayesian account of "hysteria." *Brain,* 135 (1): 3495–3512.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4: 215.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4 (11), e1000211.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, 11 (2): 127–138.

Helmholtz, H. (1866). Concerning the perceptions in general. In: *Treatise on Physiological Optics, Vol. 3* (3rd edition), trans. J. Southall. New York: Dover, 1962.

Hobson, J. A. (2009). REM sleep and dreaming: Towards a theory of protoconsciousness. *Nature Reviews. Neuroscience*, 10 (11): 803–813.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neuroscience*, 27 (12): 712–719.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427 (6971): 244–247.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121: 1013–1052.

Mumford, D. (1992). On the computational architecture of the neocortex. II. *Biological Cybernetics*, 66: 241–251.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2 (1): 79–87.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80 (1): 1–27.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, 335: 311–317.

Bodily Self, Affect, Consciousness, and the Cortex

Commentary by Vittorio Gallese (Parma)

Mark Solms's hypothesis holds that two main body representations are housed in the brain: the sensorimotor body and the autonomic body. These two body representations would be associated with two different types of consciousness: cognitive consciousness and affective consciousness, respectively. According to Solms, cognitive consciousness is secondary and depends on the primary, brainstem-located, affective consciousness. The consequence of this is that Freud's id would be conscious, while the ego would be unconscious. In my commentary, while praising Solms for his emphasis on the inseparable relation between affect and consciousness, I challenge his rigidly dichotomous account of consciousness. In so doing, I vindicate the role played by the cortex and, in particular, the cortical motor system in generating the varieties of phenomenal self-awareness we entertain.

Keywords: affect; bodily self; cortical motor system; embodied simulation; feelings; neocortex

With "The Conscious Id," Mark Solms puts forward a thought-provoking, daring hypothesis. Solms builds upon the "affective revolution" (of which he is himself a protagonist—see Solms & Panksepp, 2012) brought forward by Jaak Panksepp's affective neuroscience (Panksepp, 1998a, 1998b) and by Antonio Damasio's enlightening studies and theorization on the neural basis of human feelings (Damasio, 1999, 2010; Damasio & Carvalho, 2013). Solms's hypothesis relies on empirical evidence and has the merit of attempting to weave together the body, the nature of consciousness, its affective connotations, the role of archaic subcortical brain structures, and the bearing of all this on psychoanalytic theory and clinical practice.

The gist of Solms' proposal is the following: (1) We can distinguish two main body representations in the brain: the sensorimotor body and the autonomic body. (2) These two body representations are associated with two different types of consciousness: cognitive consciousness and affective consciousness, respectively. (3) The first type of consciousness—cognitive consciousness—is secondary and depends on the primary,

Vittorio Gallese: Department of Neuroscience, Unit of Physiology, University of Parma, Parma, Italy.

Acknowledgment: This work was supported by the EU grant TESIS.

brainstem-located, affective consciousness. (4) From this, radical consequences ensue for psychoanalytic metapsychology: Freud's id would be conscious, while the ego would be unconscious.

In this commentary, for the sake of space, I focus only on the first three points of Solms's proposal. In particular, I discuss his—in my opinion problematic—rigidly dichotomous account of consciousness, maintaining that affective consciousness is entirely subcortical, while the cortically located sensorimotor body maps would generate at best a secondary and derivative cognitive consciousness.

First of all, I would like to state that Solms's proposal undoubtedly offers a fresh view of brainstem nuclei, long conceived of as mere "switches" of neocortical activity. Brainstem nuclei and, more generally, subcortical limbic structures not only modulate cortical activity, but also contribute to the affective quality of what is being processed at the cortical level. As aptly put by Solms, "Conscious states are inherently *affective.*"

Furthermore, I very much agree with Solms when he points out that emotions are "peremptory forms of motor discharge." Solms correctly emphasizes that the distinctive feature of affective consciousness consists of the pleasure–unpleasure series, motorically expressed as approaching/withdrawal behaviors.

Indeed, a few years ago Thomas Metzinger and I argued in a similar vein that the prehistory of representational goal-states is likely to be found in the reward system, just because reward is the payoff of the self-organizing principles functionally governing and internally modeling the organization of the living body (Gallese & Metzinger, 2003; Metzinger & Gallese, 2003). Living organisms are endowed with drives pushing them toward homeostasis. Reward systems are necessary to tell the organism that it is doing right, that it is achieving a good level of integration. Thus, reward systems can be conceived of as generating a representational content of a nonsymbolic kind: the internal value assigned to a certain state. In one of those papers we wrote: "A conscious representation of value, as, for instance, expressed in a subjectively experienced emotional state, has the additional functional advantage of making survival value-related information globally available for the selective and flexible control of action, attention, and cognition within a virtual window of presence. It makes this information accessible to many different processing systems at the same time" (Gallese & Metzinger, 2003, p. 370).

The relation between goal-state representations and reward also plays a crucial role in cognitive development. Very early on, infants learn to rely on external causes for activating the reward system. Positive reactions (or their lack) to infants' behavior induced in adult caregivers provide very useful cues about how to act in a given context. Around 6 months of age, infants visually "check back" to the mother's emotional reaction in order to disambiguate how to react to certain events. Such a phenomenon is commonly designated as social referencing. The evaluation of the emotional signs of adults' reactions brings about the consolidation (or the inhibition) of a given goal-state representation.

The evolutionarily most ancient affect-related systems Solms describes in his Target Article not only provide emotional color to our behavior, but also likely provide basic and adaptive descriptions of objects[1] such as "edible," "not edible," "dangerous," "pleasurable," etc. (see Gallese, 2000). The implications for psychoanalysis could not be more obvious.

What I would like to challenge here is Solms's idea that phenomenal selfhood is the *exclusive* outcome of the upper-brainstem nuclei and of the limbic system. This view, on the one hand, while correctly criticizing the dominant corticocentric view of affective consciousness, downplays too much the role played by the neocortex in a variety of aspects of conscious life. I think that being a self whose experience of encounters with the world is constantly guided by the feelings such encounters evoke is inconceivable without the crucial role played by the neocortex. On the other hand, Solms's proposal betrays the neglect of the major role the cortical motor system plays in several aspects of consciousness, such as phenomenal body ownership and phenomenal agency (for a thorough discussion of these aspects, see Gallese, 2007; Gallese & Sinigaglia, 2010, 2011a). Briefly, it has been proposed that there is a sense of body that is enactive in nature, enabling the capture of the most primitive sense of self as bodily self. According to our perspective, the body is primarily given to us as a "source" or "power" for action—that is, as the variety of motor potentialities defining our interaction with the world we inhabit. Such primitive sense of self as bodily self is conceived of as being antecedent to the distinction between consciousness of agency and consciousness of ownership. Empirical evidence shows that the cortical motor system plays an important role in generating such a sense of bodily self (see Ferri, Frassinetti, Ardizzi, Costantini, & Gallese, 2012).

We recently addressed with an fMRI study the issue of how affect and action bind within the neocortex of healthy young participants (Ferri et al., 2013). This study shows how the emotion dynamically expressed by the face of an observed agent (happiness, anger, or neutral) modulates cortical circuits activated during the perception of her or his grasping action. As control stimuli, participants observed either the same agent's face expressing an emotion, or the agent's body performing the same grasping actions with no visible face.

The trick was that the observed grasping actions were identical in all stimuli. What changed was the absence/combination of concurrent facial expressions of positive and negative emotions of the agent. Our results show that the observation of an action embedded in the emotional context constituted by the observed agent's facial expression, when compared with the observation of the same action embedded in a neutral context, elicits higher neural responses at the level of motor frontal cortices, and of temporal and occipital cortices, bilaterally.

In particular, observing actions embedded in the context of anger, but not happiness, compared with a neutral context, elicits stronger activity in motor-related cortical areas, such as the precentral gyrus and the inferior frontal gyrus, and the presupplementary motor area (pre-SMA)—all regions playing a central role in motor control. Results suggest that the observed dynamic facial expression of anger appears to modulate the embodied simulation of the observed action. The angry context is combined with the motor representation of the observed action at the level of the cortical motor system. This triggers an immediate, context-modulated embodied simulation from the observer.

The pre-SMA plays a central role in the control of motor behavior. Its higher activation for "angry" than for "neutral action" (Ferri et al., 2013; see also Oliveri et al., 2003) can be interpreted in the light of the role the pre-SMA plays in the shaping of self-initiated actions. One could speculate that the negative emotional context connotes the perceived action as potentially threatening and, hence, evokes in the observer the embodied simulation of her or his potential motor reaction. This integration process, taking place at the level of the neocortex, probably contributes to the building of the immediate ascription of the emotional intention associated with the observed action (see Ferri et al., 2013; Gallese & Sinigaglia, 2011b).

In view of such evidence, I think we should be very careful before assuming a rigid dividing line between cognitive and affective consciousness. As recently emphasized by Damasio (Damasio & Carvalho, 2013), "feelings are likely to arise from maps of body states"; thus, "it is sensible to focus the search for neural substrates of feelings on the regions exhibiting topographically organized somatic maps" (p. 146). They conclude that ". . . the most prominent system level candidates for neural substrates of feelings can be found on two distinct phylogenetic levels: the more primitive level of the brainstem (specifically, the parabrachial nucleus, the nucleus tractus solitarius, the periaqueductal grey and the deep layers of the superior colliculus) and the more recently evolved cerebral cortex (specifically, the insula, SI and SII [somatosensory I and II])" (p. 146). I would certainly add cortical motor areas to this list.

Three levels of selfhood have been identified from a phenomenological point of view (see Parnas, 2000, 2003). The first one consists of the implicit awareness that this is "my" experience. Such pre-reflective level of selfhood is sometimes referred to as the "basic" or "minimal" self, or as "ipseity." The second level consists of the more explicit awareness of self as an invariant subject of experience and action. Such a reflective level of self-awareness presupposes the "minimal" self. Finally, there is the social or narrative self, which refers to personality, habits, style, and other characteristics of an individual.

I agree with Solms when he says that our conscious thinking is "*constantly accompanied by affect*." However, on the basis of the currently available neuroscientific evidence, I am not convinced that a phenomenal first-person perspective can be exclusively explained by affect and its subcortical underpinnings. For the very same reasons, I am even less convinced that the "cortex is nothing but random-access memory" and that Freud's "bodily ego" can *only* become conscious "when cathected by the id." Luckily enough, all of these issues can be empirically investigated by cognitive neuroscience.

REFERENCES

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt Brace.

Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon.

Damasio, A., & Carvalho, G. B. (2013) The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews. Neuroscience*, 14: 143–152.

Ferri, F., Ebisch, S. J., Costantini, M., Salone, A., Arciero, G., Mazzola, V., et al. (2013). Binding action and emotion in social understanding. *PLoS One*, 8 (1), e54091.

Ferri, F., Frassinetti, F., Ardizzi, M., Costantini, M., & Gallese, V. (2012). A sensorimotor network for the bodily self. *Journal of Cognitive Neuroscience*, 24 (7): 1584–1595.

Gallese, V. (2000) The inner sense of action: Agency and motor representations. *Journal of Consciousness Studies*, 7: 23–40.

Gallese, V. (2007). The "conscious" dorsal stream: Embodied simulation and its role in space and action conscious awareness. *PSYCHE*. http://psyche.cs.monash.edu.au

Gallese, V., & Metzinger, T. (2003). Motor ontology: The repre-

sentational reality of goals, actions, and selves. *Philosophical Psychology*, 13 (3): 365–388.
Gallese V., & Sinigaglia C. (2010). The bodily self as power for action. *Neuropsychologia*, 48: 746–755.
Gallese, V., & Sinigaglia, C. (2011a). How the body in action shapes the self. *Journal of Consciousness Studies,* 18 (7–8): 117–143.
Gallese V., & Sinigaglia, C. (2011b). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15 (11): 512–519.
Metzinger, T., & Gallese, V. (2003). The emergence of a shared action ontology: Building blocks for a theory. *Consciousness and Cognition*, 12: 549–571.
Oliveri, M., Babiloni, C., Filippi, M. M., Caltagirone, C., Babiloni, F., et al. (2003). Influence of the supplementary motor area on primary motor cortex excitability during movements triggered by neutral or emotionally unpleasant visual cues. *Experimental Brain Research*, 149: 214–221.
Panksepp, J. (1998a). *Affective Neuroscience*. New York: Oxford University Press.
Panksepp, J. (1998b). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the SELF. *Journal of Consciousness Studies*, 5: 566–582.
Parnas, J. (2000). The self and intentionality in the pre-psychotic stages of schizophrenia: A phenomenological study. In: *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience,* ed. D. Zahavi. Amsterdam: John Benjamins, pp. 115–148.
Parnas, J. (2003). Self and schizophrenia: A phenomenological perspective. In *The Self in Neuroscience and Psychiatry*, ed. T. Kircher & A. David. Cambridge: Cambridge University Press, pp. 127–141.
Solms, M., & Panksepp, J. (2012). The "Id" knows more than the "Ego" admits: Neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science,* 2: 147–175.

Conflict Creates an Unconscious Id

Commentary by Jim Hopkins (London)

In relating Mark Solms's framework of subcortical id and cortical ego to psychoanalysis, we should bear in mind the particular circumstances of human emotional conflict. Freud explicated this in terms of the superego, which was the first internal object and was also used in relating emotional conflict in the individual to violent group conflict. In describing conflict, we specify conditions in which the translation of subcortical oscillation into conscious action-directing representation that Solms describes should break down, and as psychoanalysis describes using the concept of an unconscious id. Considering attachment suggests likewise. This shows both that Solms's framework matches psychoanalysis and that we should revise the latter not by regarding the id as conscious but by emphasizing that the unconscious and the id are created via the joint generation of motivation and consciousness in the evolved context of human emotional and group conflict.

Keywords: attachment; compromise-formation; conflict; projection; RAGE; superego

In a creative and far-reaching argument, Mark Solms maps the id to the basic subcortical mechanisms of motivation and the ego to thalamocortical systems, transforming "fleeting, wavelike" activations from these into temporally ordered cortical representations informing bodily activity. This translation of id by ego enables our bodies to discharge the functions of the subcortical mechanisms in the cortically represented environment into which we have evolved. Since the mechanisms that generate emotion also generate consciousness, he concludes that the id is both conscious and the source of consciousness for the cortical action-directing mechanisms as well. This contradicts the psychoanalytic account, according to which the motivations and representations in the id are unconscious. What revisions are required?

To answer this we must consider conflict. Many paradigms of psychoanalysis are also paradigms of disordering emotional conflicts that psychoanalysis seeks to mitigate. Such conflict might well originate in affordance competition (Cisek & Kalaska, 2010), the psychoanalytic cases marking failure of winner-take-all in incoherent activation of multiple competitors. (Dreams show comparable multiple activation but serve homeostatic optimization during sleep.) The activations associated with the ego, superego, and id are a further special case, hypothesized by virtue of

Jim Hopkins: Research Department of Clinical and Health Psychology, University College London; Department of Philosophy, Kings College London, London, U.K.

Freud's growing realization (explicit in his notes on the Rat Man; Freud, 1909) that conflict characteristically involves not just the activation of multiple conflicting motives, but also that of accompanying constellations relating to conflicting parts or aspects of the self.

The most familiar (making use of Panksepp's terminology) is that of moralistic superego RAGE punishing the self/ego for activation of RAGE in the id. This can be seen in the Rat Man's anxiety and guilt for the unconscious RAGE at his father expressed in his phantasy of paternal torture. Freud reasonably regarded this as an evolved mechanism for controlling aggression via guilt, functioning to secure family harmony and facilitating ingroup cohesion for outgroup aggression, because internal identification with the superego locates (projects) badness outside, both into other individuals and into outgroups, against whom moralistic RAGE is therefore amplified as mandatory.

We see this in the psychodynamic description of countless examples of the *good us vs. bad them* pattern that underpins group conflict—in religious conflicts and wars, the facilitation of the Holocaust by the idealization of Hitler, current demonization of Muslims and idealization of militarism in the United States, and so on *ad finem nostrum*. The same holds for individual good-self/bad-other relations, as was first exhibited in full detail in the working of Freud's moralistic superego as collector of instances of his own medical mispractice, brought against himself—and threatening talionic death for his own daughter—in his associations to his dream of "Irma's injection" (Freud, 1900, p. 107). The wish-fulfillment in the first specimen of psychoanalysis was obtained by Freud's evading the guilt-producing daughter-menacing superego whose nocturnal working his associations revealed, by projecting mispractice into Otto[1] so as to vent RAGE on Otto in taking the superego role for himself.

The opposite phenomenon—conflict-relieving projection of the superego—is illustrated in the common passage from depression to schizophrenia vividly described after superego–self-aggression in Elyn Saks' account of her descent into psychosis (Saks, 2007). Similar superego/id dialectics are enacted in countless psychoanalytic examples, including children's play, as described by Melanie Klein (1929).[2] As well as explicating many phenomena, the superego–ego–id framework introduced subpersonal functional units whose personification embodies an insight fundamental for neuropsychoanalysis—namely, that, in our social species, subpersonal functional/causal roles for regulating emotion are cortically discharged by neural/mental representations of human beings themselves: by the superego and by other internal objects.

Freud's various accounts of drives (one involving RAGE often opposed to others, as in death/life) were framed to explicate such conflict. This is no accident, since such conflict specifies conditions in which the transformation of subcortical activation into conscious representation breaks down. Pre-superego conflict involves activation of RAGE together with other basics opposed to it toward a single object—which is inevitable when all urgent activations must be directed at a nursing mother. But, in this, the cortical ego must deal with subcortical activations requiring transformation into *contradictory* representations in behavior and consciousness. In such a situation, we should expect the outcomes that psychoanalysis has repeatedly found: *either* one of the potentially conflicting id activations is denied translation, as in repression; *or*, *failing this*, both get partial but incoherent translation, as in compromise-formations such as Freudian symptoms or the contradictory/incoherent behavior of infants classified as attachment-disorganized at age 12 months.

Frequent conflicting outcomes would be probable if subcortical oscillators superpose conflicting imperatives in a single waveform for cortical resonators to unfold and broadcast as the *changing* sequential patterns required for emotionally complex and changing situations. Here, subcortical superposition would explain how fleeting waves manage the required information, how integration of conflicting signals would be the ego's normal task, and how incoherent cortical activation would be probable in difficult conditions. Such conflict would naturally activate past conflict-engendering repression-causing representations from the earliest parental imagos and would also be expressed in concurrent phantasies—conscious and unconscious—involving them. Such might be exemplified in a projected form in the Rat Man's transference phantasies of Freud's mother dead with swords stuck though her breast, as well as his imagining Freud as a murderous beast of prey (projected moralistic RAGE-filled devouring and body-invading oral superego) that would fall on him to search out what was evil in him.

How pervasive is such conflict? Psychoanalysis apart, it appears that mother–infant attachment—which Watt and Panksepp (2009, p. 93) describe as setting the "massive regulatory-lynchpin system of the human brain" exercising "primary influence over the prototype systems below"—must involve particularly powerful conflicts about RAGE. Attachment inhibits

[1] Dr. Oskar Rie, the "Irma" family's pediatrician, whom Freud referred to as "Otto."

[2] Examples are discussed together with a Bayesian account of the superego and repression compatible with Solms's framework in Hopkins (2012) with references to material in this commentary.

RAGE and FEAR that might conflict with SEEKING and CARE between the attached. Central to CARE, in turn, is proximity, and this is regulated by separation-distress/PANIC/GRIEF (SPG), which is itself a cause of RAGE and FEAR, for these too are marshaled in maintaining proximity—as seen in the RAGE at separation from mother and FEAR of strangers (the first of many *bad them*) appearing in human attachment at age 7–8 months—and also because, as with FEAR of strangers, they serve to prepare the potentially separated for the worst.

Powerful frequent activations of this SPG/RAGE/FEAR triad seem essential in early infancy, as their expression in infants' uniquely penetrating, anxiety-arousing, and action-compelling cries are the main means by which those in urgent need can coerce the CARE that is a continual matter of life and death for them. The importance of this triad appears in its determination of attachment classification as secure, or again as insecure (avoidant, ambivalent-resistant, or disorganized). In the "strange-situation" procedure, mother and infant settle in an unfamiliar (potentially FEAR-producing) situation, and the infant begins SEEKING and PLAY. A stranger (potential *bad other*) enters and approaches the infant, activating the infant's recently lessened FEAR of strangers. After this, the mother, following procedure, leaves the infant alone by itself, activating the infant's recently lessened RAGE at maternal separation and therewith the infant's SPG in this context of FEAR and RAGE. After reunion, the mother activates the triad still further, leaving the infant again, this time at the mercy of the stranger, who originally activated FEAR. This activates RAGE again at the mother for doing all this, and despite the infant's previous protesting RAGE.

Classification depends upon how the infant copes, particularly with RAGE at the mother when she returns offering renewed CARE, whose enjoyment conflicts with RAGE. (The situation, that is, has now manipulated the triad to activate the central evolutionarily significant conflict—RAGE and FEAR in opposition to the SEEKING of CARE—that attachment itself serves to contain.) In this situation secure infants show conflict but resolve it to return CARE, and thence to SEEKING in PLAY. Disorganized infants, by contrast, remain in conflict, apparently involving activated RAGE and FEAR, and they are confused in feeling, and contradictory and incoherent in behavior, and can make no comparable return.

This is how it would be if the secure infants' cortical egos had learned, as the disorganized infants' had not, to regulate potentially conflicting subcortical activations so as to translate them for action situations demanding rapid changes in the direction of RAGE. The importance of *early* conflict, moreover, is further indicated by the fact that disorganized attachment can be predicted from emotional dis-coordination between mother and infant at age 4 months (the time of coming to represent the mother as a whole object that Klein (1952) thought initiated her paranoid-schizoid→depressive transition). And the formative role of RAGE in the core conflict of attachment seems to be shown in the disposition of disorganized infants later to become violently controlling in respect of CARE.

All this fits with the psychoanalytic account in terms of superego, ego, and id, including Freud's idea that early conflict-generating imagos play a key role in development, and also with Klein's hypothesized transition. We should therefore include the possibility that after the mother is singled out between age 4 and 5 months as a unique and irreplaceable source of CARE (thereby prompting the separation-RAGE and stranger-FEAR apparent at age 7–8 months), the infant's experience of this might promote an infantile version of CARE in return. In this, the GRIEF of SPG might become more distinctly human grief, so that the infant would now curb RAGE out of something like gratitude and concern that RAGE might damage this source of CARE or provoke its loss.

Solms stresses that psychoanalytic claims about the id were framed to explain clinical data and remain answerable to them, and he aims to provide an account of what is made unconscious and how. Together with the above, this should encourage us to depart from his wording in integrating his valuable account. We have seen that the id is part of an explanatory structure introduced for, and serving, genuine explanatory purposes, and also that this is not just consistent with, but apparently a consequence of, the neuroscientific framework that Solms advocates. Moreover, the main limitation of this structural theory—that the superego is only the first-discovered of a variety of differing internal objects in distinctive causal-functional roles—is consistent with expanding it within his framework. The introduction of the superego as an internal object inhibiting, projecting, and amplifying RAGE so as to subserve ingroup cohesion for outgroup conflict was also the introduction of internalized object-relations as a potential theoretical framework. An ego that creates working representations of objects will also organize the whole theatre of motivation- and conscious-regulating prototypes of emotional relationships embodied in internal objects, which are structured by within-attachment conflicts and so subject to repression.

Given such convergence we should not integrate Solms's framework by describing the id as conscious.

Assuming confirmation, we should just accept the framework as the current best account of the joint generation of motivation and consciousness and, in light of this, make explicit that the unconscious and the id are not intrinsically linked with the motive/consciousness mechanisms. Rather, they are to be seen, as indicated above, as *consequences* of the working of these mechanisms—and especially of the burden put on the id-deciphering infantile ego—in our uniquely social (and so moralistically RAGE-limiting) and also uniquely and lethally group-competitive (and so moralistically RAGE-projecting and RAGE-amplifying) and therefore uniquely conflicted mammalian species.

REFERENCES

Cisek, P., & Kalaska, J. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33: 269–298.

Freud, S. (1900). *The Interpretation of Dreams*. *Standard Edition*, 4/5.

Freud, S. (1909). Notes upon a case of obsessional neurosis [Original record of the case]. *Standard Edition*, 10.

Hopkins, J. (2012). Psychoanalysis representation and neuroscience: The Freudian unconscious and the Bayesian brain. In: *From the Couch to the Lab: Psychoanalysis, Neuroscience and Cognitive Psychology in Dialogue*, ed. A. Fotopoulou, D. Pfaff, & M. Conway. Oxford: Oxford University Press.

Klein, M. (1929). Personification in the play of children. In: *The Writings of Melanie Klein, Vol. 1: Love, Guilt and Reparation and Other Works*. London: Hogarth Press, 1975.

Klein, M. (1952). Some theoretical conclusions regarding the emotional life of the infant. In: *The Writings of Melanie Klein, Vol. 8: Envy and Gratitude and Other Works*. London: Hogarth Press, 1975.

Saks, E. R. (2007). *The Center Cannot Hold: My Journey Through Madness.* New York: Hyperion.

Watt, D., & Panksepp, J. (2009). Depression: An evolutionarily conserved mechanism? *Neuropsychoanalysis*, 11 (1): 87–109.

Conscious Id or Unconscious Id or Both: An Attempt at "Self"-Help

Commentary by Luba Kessler (Long Island City, NY)

The concept of the "self" is proposed as a potential mediator in the questions raised by Mark Solms regarding bodily representations in the brain/mind and their relationship to consciousness. The insights of mid-twentieth-century psychoanalytic researchers and of Kohut's self-psychology are used as a springboard for correlations with modern-day findings of affective neuroscience. Their addition may shift the enduring Freudian metapsychology into a more workable relationship to the neurosciences, and vice versa.

Keywords: insula; metapsychology; re-presentation; self; self-states

In bringing to bear neuroscientific findings on psychoanalytic metapsychology, the Target Article both affirms the utility of the Freudian model and calls for its reform. Its basic proposition is a reassignment of the intrinsic capacity for consciousness from the Freudian ego to the id. Mark Solms's invitation is based on the findings of affective neuroscience that the subcortical repository of the value-laden affective signals from the interior of the body—necessary for the regulation of homeostasis—in the upper brainstem is what constitutes the basic, *sine qua non*, layer of consciousness: "When endogenous consciousness is obliterated, exteroceptive awareness is obliterated too." This contradicts the basic psychoanalytic proposition of the unconscious nature of bodily-derived instinctual id drives. Further along in the article, Solms similarly undermines the basic psychoanalytic precept of the ego as the seat of consciousness. He points out, interestingly, the ego's tendency to economize mental activity by automatization of psychic solutions into implicit procedures, relieving it of the need of consciousness. What follows is a view of neurotic illness as captivity to maladaptive premature automatization and of the therapeutic action of psychoanalysis as the reawakening of episodic associative links to arrive at new executive programs for more adaptive automatization. This gives support to Loewald's (1960) therapeutic action of a "new" object as one facilitating new episodic experience within the psychoanalytic situation.

Luba Kessler: Institute for Psychoanalytic Education, affiliated with the New York University Medical School, New York, NY, U.S.A.

It is a pleasure to follow Solms as he tests Freudian thought against the insights derived from neuroscience. It is a live dialogue between the two domains of inquiry in which each is sharpened by the use of the methodologies of the other. It bears out the enthusiasm of his closing words announcing that ours is a Golden Age of neuroscience on its way to becoming a truly mental science.

"The Conscious Id" seemed convincing to me in its conclusions and proposal for a reformed metapsychological treatment of consciousness. However, when Solms introduces into his article the concept of "self," he opens up a space for further questions and considerations.

The reference first arrives at the end of Section 3, on the corticocentric fallacy of viewing consciousness as a cortical function, when he states: "He [Bud Craig (2009)] equates this cortical region [of posterior insula] with the body-as-subject, the primary sentient 'self'—precisely the function that I have attributed, on the basis of a different research tradition, to the upper brainstem and limbic system." However, what Solms had attributed earlier in the article to the upper brainstem was the representation of the "internal body" and its subjective "state" as an aspect of consciousness, not a "self." Echoing Freud's "A note upon the "Mystic Writing-pad" (1925), he says in that context: "We may picture this aspect of consciousness as the page upon which external perceptions are inscribed." No "self" is considered as either a metapsychological or neuroscientific concept in its own right here.

Does it matter? It certainly seems to matter in the context of Solms's own argument against the corticocentricity of consciousness. As an example, he cites unequivocal evidence of affective consciousness in hydranencephalic children (Merker, 2007). However, when he appears to equate their apparent manifestations of affective consciousness in the absence of a cortex with evidence of a primary (affective) "self," he asserts equivalence between the two concepts without making the case for it. Consciousness appears to be used synonymously with "self." But are they one and the same?

Similarly, Solms's argument against Craig's (2009) research findings of the existence of a cortical projection zone for the internal body in the posterior insula provides a further opportunity to trace this possible conflation of consciousness with "self." Solms cites Damasio's (Damasio, Damasio, & Tranel, 2012) finding that in a case of total bilateral obliteration of insula by herpes simplex encephalitis, a patient nevertheless clearly manifested a well-preserved and robust selfhood—a finding contradicting Craig's assertions. What this argument seems to overlook, though, is that what the experiment as described tried to contest but did not controvert was a previously established selfhood, encoded as it was in a multitude of self-representations built on the primary sentience presumably existing before the damage to the insula occurred. Research would need to establish what the nature of the effects of an *early* insult to insula might have on that primary bodily self-sentience and its developmental consequences in order to learn the relationship between the affective "self" and affective consciousness.

So, if there is a distinction between them, how might we think about what makes them different? It seems to me that it is possible to think of that difference in terms of the representational activity of the brain/mind: the *global affective consciousness* expressing the subjective state at the subcortical level gathered for the next level of re-presentation in the cortex as a *primary bodily "self."* The insula would seem to be the prime candidate for such a function. Its embedded location at the base of a deep cortical fold brings it, in fact, very close to the neighboring upper-brainstem topography, where affective components for consciousness are generated. Indeed, on macroscopic viewing of the brain, it is like a cortical island (hence its name?), surrounded as it is by the subcortical tissue. Because of that, the insula seems to be a candidate *par excellence* for the primal global cortical representation—a *"self"*—of its subcortically gathered internal affective state. This insular representation of the self is not of a nature of later self-representations such as develop further along in the course of accruing interactions of the "self" with the objects in the world; rather, it may indeed be their first basic layer. Is it possible that it is where Damasio's (2010) "protoself" accrues its emergence? Is it the seat of Freud's "mystic writing-pad" (1925), on which further experiences are inscribed? In keeping with Solms's ideas of the brain's inclination toward its own economical thrift, it would seem of considerable adaptive evolutionary purpose to have a structure that gathers all the affective signals from the state of the body's interior into a representation of a *"self"* capable of participating in the psychophysiological regulation of homeostasis.

If these ideas have any merit, credit goes to the conceptualizations of the self-psychological tradition of psychoanalysis. By inviting attention to *states of the self* as a particular domain of the mind, Kohut's (1971) self-psychology extended psychoanalytic interest to primal experiences of the "self" even before the advent of affective neuroscience. It began to give its own quasi-metapsychological dimensions to the states of self: there could be strong or feeble, cohesive or fragmenting, robust or disintegrating selves. Classical psychoanalysis and its metapsychology has had a hard time including consideration of such formulations into its own theoretical body, even while acknowledging some of their clinical resonances. This may be because the global experiential quality of these self-states ill-fits into the body of psychoanalytic thought, which has viewed the "self"—as Solms argues—as an ego-cortical product of objectified representations. Even still, as things stand, considerations of such states generally tend to fall under the purview of attachment and infant research—at a considerable remove from psychoanalytic metapsychology.

This was not always the case within psychoanalysis. Hartmann's (1950; Hartmann, Kris, & Lowenstein, 1946) and Jacobson's (1964) ideas about the undifferentiated psychophysiological *ur*-self considered, at least theoretically, the existence of global states awash in the primal pleasure–unpleasure affects of the organism's *milieu interieur*, as did Mahler, Pine, and Bergman (1975) in postulating the "autistic" phase at the start of the separation–individuation process of development. Spitz (1965) brilliantly intuited the "primal cavity" world of global coenesthetic perception of the earliest nursing situation, echoed in the altered consciousness states of the "Isakower phenomena" (Isakower, 1938). Bertram Lewin (1946) postulated the primal "dream screen," recently revisited by Lehtonen et al.'s (2006) paper, "Nascent Body Ego," about the ascendance of theta-wave activity in mental development. These traditions seem to have fallen out of the psychoanalytic library of ideas, and it may be that, as in Lehtonen, they might find a second-coming through the insights of neuropsychoanalysis.

It would seem that these metapsychological intuitions of mid-twentieth-century psychoanalytic researchers fit well into the articulations of affective consciousness within neuroscience. Perhaps along with Solms they would assume that subcortical affective subjectivity presumed an "*anlage*" [primordium; developmental germ] of the primal "self." They did not, however, say this, largely taking cover in the concepts of primary narcissism or the undifferentiated psychophysiological self. On the other hand, perhaps they would take up the idea of the insula as a particular way-station in the process of the substantiation of the representational activity leading to the development of "self."

What does it portend for Solms's propositions about the assignation of consciousness to the metapsychological concepts of the id or ego, one way or the other? If consciousness is a property that dawns in the first global representation of an affective state in the upper brainstem, does it reside in the id? If consciousness is what becomes substantiated through the re-presentation of that affective state as a primal "self" in the cortical structure of the insula, does it mean that it resides in the ego?

It seems to me that a resolution of this question depends on how we view the relationship of *consciousness-as-property* to *consciousness-as-system*. I believe that Solms argues very convincingly that it is a property that emerges through the bottom-up communication of affectively laden signals from the interior of the body, accruing new representations in their journey to cortical distribution and reciprocal top-down regulation. In the period of transitioning from neurology to psychology in the 1890s, Freud wrote to Fliess: "I should like to emphasize the fact that the successive registrations represent the psychic achievement of successive epochs of life. At the boundary between two such epochs a translation of the psychic material must take place" (Freud, 1896, p. 208). Psychological development is a matter of successive representations. One hundred years later, Edelman (1987, 1992) would echo related ideas in the neuroscientific language of his neural Darwinian theory of neuronal group selection, suggesting that new kinds of brain morphology come about through the process of somatic selection driven by changing levels of homeostatic demands in the course of development. Translated into self-psychological terms, the morphology of the brainstem serving the self-state of affective consciousness morphs selectively into neuronal grouping of a different kind at the next station of homeostatic activity, to produce a dawning conscious representation of "self."

If id is where the bodily imperatives originate, then it is the source of consciousness, whereas the ego is the product of its successive re-presentations. In this view, the mind is viewed as an apparatus for (interacting) representational layering, as per Freud and Edelman. If, on the other hand, in consequence of the complexities of its relationship with external objects—so-called reality—the inevitable "vicissitudes" force the intrinsic id drives out of the linear representational developmental dynamics into the pockets of primary or secondary repression (Freud, 1915), then the id is the source for what remains, or becomes, unconscious. This is the mind as a censoring apparatus, safeguarding the negotiations for optimal homeostasis of the subject in the world. Here, the emphasis is on the interaction—and conflict—between Freud's pleasure and reality principles of mental functioning (Freud, 1911).

It may be that the "self" is where the two domains of psychobiological imperatives converge: one charged with representational activity of its interior state (i.e., narcissism), the other with stemming this narcissistic expansiveness at its confrontation with the reality of the object-world. The deeply inverted cortex of the insula may be where the agenda of the expansive subcortical growth of the pleasure-seeking self-state from within elaborates a rudimentary self-representation required for the task of differentiating itself from the objects it has to negotiate with to protect its homeostatic security in the external world—that is, reality testing. In that case, the id is at once the fount of consciousness and the unconscious repository of some of its contents following an encounter with the "self" and its agenda of self-preservation. How we think about those assignations within the id may depend on the question: What do we mean by Id? (Hayman, 1969)

It is interesting in this regard to learn from a finding by Gallese (Ebisch et al., 2011), reported by Solms, about the ability of the mirror neurons to differentiate between "me" and "not-me" movements, in which reportedly the suppression of insular activation appears to play a role. These are intriguing findings, which feed the curiosity about the insula's participation in the formation of the rudimentary "self." Admixing poetic license, one might envision its role as being akin to an internal mirror where the first reflection of the subjective state creates a "self" as the image of its own dawning representation.

In summary, by engaging the insights of psychoanalytic metapsychology and neuroscience, Solms prods us to think anew about the accepted wisdom of our learning in these respective disciplines. This discussion has taken up his invitation to review the metapsychological formulations by stretching them so as to include the perspectives of the psychology—and perhaps the neuroscience as well—of the self. The hope is that the phenomenology of the self, as viewed from both a psychological and a neuroscientific perspective, would provide some missing links in the understanding of the workings of the mind. It suggests that adding the insights of the mid-twentieth-century psychoanalytic thinkers into the development of the self and their subsequent reformulations by Kohutian self-psychology might advance Solms's daring inquiry into the salience of consciousness for the understanding of human psychology.

REFERENCES

Craig, A.D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, 10: 59–70.

Damasio, A. (2010). *Self Comes to Mind.* New York: Pantheon.

Damasio, A., Damasio, H. & Tranel, D. (2012). Persistence of feeling and sentience after bilateral damage of the insula. *Cerebral Cortex* [Epub ahead of print]. doi:10.1093/cercor/bhs077

Ebisch, S. J. H., Ferri, F., Salone, A., Perrucci, M. G., D'Amico, L., Ferro, F. M., et al. (2011). Differential involvement of somatosensory and interoceptive cortices during the observation of affective touch. *Journal of Cognitive Neuroscience*, 23 (7), 1808–1822.

Edelman, G. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection.* New York: Basic Books.

Edelman, G. (1992). *Bright Air, Brilliant Fire.* New York: Basic Books.

Freud, S. (1896). Letter from Freud to Fliess, December 6, 1896. In: *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887–1904*, trans J. M. Masson. Cambridge, MA: Belknap Press, 1985, pp. 207–214.

Freud, S. (1911). Formulations on the two principles of mental functioning. *Standard Edition*, 12: 213–226.

Freud, S. (1915). Instincts and their vicissitudes. *Standard Edition,* 14: 117–140.

Freud, S. (1925). A note upon the "mystic writing-pad." *Standard Edition*, 16: 227–232.

Hartmann, H., Kris, E., & Loewenstein, R. M. (1946). Comments on the formation of psychic structure. *Psychoanalytic Study of the Child,* 2.

Hartmann, H. (1950). Comments on the psychoanalytic theory of the ego. *Psychoanalytic Study of the Child,* 5

Hayman, A. (1969). What do we mean by Id? *Journal of the American Psychoanalytic Association,* 17: 353–380.

Isakower, O. (1938). A contribution to the pathopsychology of phenomena associated with falling asleep. *International Journal of Psychoanalysis,* 19: 331–345.

Jacobson, E. (1964). *The Self and the Object World.* New York: International Universities Press.

Kohut, H. (1971). *The Analysis of the Self.* Chicago: University of Chicago Press.

Lehtonen, J., Partanen, J., Purhonen, M., Valkonen-Korhonenen, M., Kononen, M., Saarikoski, S., et al. (2006). Nascent body ego: Metapsychological and neurophysiological aspects. *International Journal of Psychoanalysis*, 23: 438–445.

Lewin, B. (1946). Sleep, the mouth, and the dream screen. *Psychoanalytic Quarterly,* 15: 419–434.

Loewald, H. W. (1960). On the therapeutic action of psychoanalysis. *International Journal of Psychoanalysis,* 41:16–33.

Mahler, M., Pine, F., & Bergman, A. (1975). *The Psychological Birth of the Human Infant.* New York: Basic Books.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Science*, 30: 63–134.

Spitz, R. (1965). *The First Year of Life.* New York: International Universities Press.

Back to the Future with Mark Solms: The Isomorphism of Freud's Mental Apparatus

Commentary by Richard J. Kessler (New York)

Mark Solms has written an article that suggests that neuroscientific data regarding the brain's generation of consciousness pose a major challenge to Freud's model of the mental apparatus. In doing so, however, he has demonstrated the vitality of that model—its continuing relevance in our efforts to understand not just the mind but also how the mind and brain are related.

Keywords: hallucinatory wish; isomorphism; metapsychology; perception; primary process

Mark Solms has written a remarkable article that deserves to be a landmark in the history of psychoanalysis, but not only because it challenges long-held beliefs. Among its many accomplishments, it demonstrates the vitality and adaptability of the Freudian model of the mind, a model that can confront contradictions and change in important ways without sustaining damage to its basic integrity. Thereby, the article makes clear the continuing value of metapsychology in providing a bridge to the neurosciences. Indeed, it could be said that metapsychology is proving to be the *lingua franca* of neuropsychoanalysis itself—the very hyphen in neuro-psychoanalysis! The article makes a strong case for a remarkable resonance between the Freudian model of the mind, with its embedded, albeit insufficiently articulated model of the brain, and current understanding of crucial aspects of brain functioning. In referencing Freudian propositions about the primacy of affective experience, the role of hallucination and memory in consciousness, the process of reality testing in the learning of perception, and the concepts of the primary and secondary process, Solms finds an impressive isomorphism with modern neuroscientific findings. Along the way, he resumes Freud's quest to infuse Darwinian evolution into an understanding of Man's nature by extending it beyond psychology and more fully to neuroanatomy and neurophysiology.

To achieve all this, Solms necessarily takes us back to 1900 and the dawn of psychoanalysis, a time when the soil on the grave of the "Project" (Freud, 1895) was still fresh. In their 1986 paper, "On Psychoanalysis and Neuroscience," Solms and Saling outlined—through a review of letters written to Fleiss—the transformation of the repudiated "Project" to the model of the mental apparatus in Chapter 7 of *The Interpretation of Dreams* (1900). In this process, Freud is said to have completely revised the "Project," renamed it *metapsychology*, and then "borrowed" a theory from *On Aphasia* (1891). No wonder that the Chapter 7 model, as many have pointed out, including Strachey (1930), in his introduction to *The Interpretation of Dreams*, shares many features of Freud's neurological speculations. This is nowhere better illuminated then in Freud's monograph *On Aphasia* itself. In Freud's illustration of the difference between object and word associations (Figure 1), which will become the distinction between word- and thing-presentations at the foundation of the concepts of primary and secondary process, he says the following: "According to philosophical teaching the idea of the object contains nothing else; the appearance of a 'thing' the 'properties' of which are conveyed to us by our senses, originates only from the fact that in enumerating the sensory impressions perceived from an object, we allow for the possibility of a large series of new impression to the chain of associations (J. S. Mill). This is why the idea of the object does not appear to us as closed, and indeed hardly as closable, while the word concept appears to us as something that is closed though capable of extension" (p. 78).

Solms cites the findings of Carhart-Harris and Friston (2010) on the brain economics of bound and unbound energies as consistent with the Freudian understanding of how the unbound drive energies of the primary process of the pleasure principle come to be constrained—that is, bound by the secondary pro-

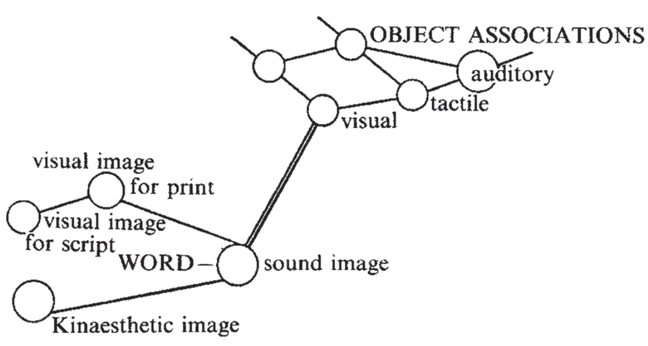

Figure 1. Diagram from *On Aphasia* that depicts the constraining effect of the word on object associations.

Richard J. Kessler: Adults & Children with Learning and Developmental Disabilities, Inc., New York, NY, U.S.A.

cesses of the reality principle. Friston's (2010) model describes in computational terms how wish (incentive salience) becomes prediction so that the brain's model of the world becomes more veridical. Freud describes the prototypical circumstance for this process:

> An infant at the breast doesn't as yet distinguish his ego from the external world as a source of sensations following in upon him. He gradually learns to do so, in response to various promptings. He must be strongly impressed by the fact that some sources of existence which he will later recognize as his bodily organs, can provide him with sensations at any moment, whereas other sources evade him from time to time—among them what he desires most of all, his mother's breast—and only reappear as a result of his screaming for help. In this way there exists for the first time set over against the ego an "object," in a form of something which exists "outside" and which only is forced to appear by a special action. [1930, pp. 65–66]

Solms's statement in Section 6, "Learning entails the establishment of associations between interoceptive drives and exteroceptive representations, guided by feelings that are generated in such encounters," might summarize this process. Implied in this, too, is that perception is learned by virtue of the gradual distinctions being established between the subjective and objective, the self and the object, and the inner and outer worlds. Not unimportant is the understanding that the world is discovered through the object.

Solms notes, however, that for all this to transpire there must be a constraint on motor discharge and tolerance of frustrated emotions so that fresh thinking and thus new learning can arise. This is an initial requirement of Freud's theory of the development of thinking.[1] The wishful hallucination, an experience-generated image of the body and the absent object, must be held in mind so as to allow working memory to find a reasoning strategy. In fact, Damasio (1994) calls the mind a cognitive process involving body images "over which you reason" (p. 196). To complete the process by which the world becomes populated with empirical objects, the Freudian model requires a renunciation of the hallucinatory wish and the differentiation of memory from current experience. Opatow (1999) describes this process: propelled by the pain of the failure of wishful hallucination to fulfill somatic needs, the infant achieves

> an inner distancing that converts the hallucinatory immediacy of sense into an explicit experiential memory ("in the past"). Furthermore, this withdrawal (or recoil) of consciousness "makes room," in awareness, for syncretic images to take on a determinate definition as distinct objects. In the rejection of the hallucinated image, consciousness negates sensations by attributing them to objects (as properties) rather than be merely assailed by them (as in hallucination). [p. 103]

As Freud (1896, pp. 278–279) declared, ". . . reality–wish-fulfillment. It is from this contrasting pair that our mental life springs."

Of course, the role of hallucination in the generation of consciousness is in itself one of Freud's most radical proposals. Intuiting this from his study of dreams, and their hallucinatory nature "counting as undisputed reality," he stated rather emphatically that "Nothing prevents us from assuming that there was a primitive state of the mental apparatus in which this path was actually traversed, in which wishing ended in hallucinating" (1900, p. 566). The central role of hallucination in perceptual consciousness, rarely recognized as a Freudian concept, has gained increasing neuroscientific credibility, to the point, as Solms notes (Section 1, footnote 4), that "What is now widely accepted is the once radical notion that perceptual consciousness is endogenously generated; exteroceptive stimuli merely constrain and sculpt what is fundamentally a hallucinatory process."[2] Remarkably, even Alan Hobson (2009), Solms's dream-theory adversary of several decades, has acknowledged the significance of dreaming (as a protoconsciousness) in the development of waking consciousness.

But what of Freud's corticocentric bias? As Solms implies, it would seem logical that once he had situated the generation of mental activity in the primitive, "instinctual" recesses of the mind, a similar consideration of the brain should follow. Apparently, despite Freud's

[1] One wonders if the motor paralysis during REM sleep serves as a physiological model for the inhibition necessary for this to take place.

[2] Changes in the dynamic relationship between internally generated mental activity and incoming sensory information produce a variety of alterations in the quality of consciousness. Impair or reduce the reception, processing, or integration of sensory information and you are likely to experience hallucinations. This occurs, for example, in psychoses like schizophrenia, sensory deprivation, experimental blindness, and anosognosia. Even the reception of poor-quality visual information—as in Charles Bonnet Syndrome and distorted proprioceptive stimuli as with weightless space-flight astronauts—can produce hallucinations (Newberg, 1994). Similarly, as a result of the degraded or degrading state of executive function, one is more likely to dream during NREM Stage 1 sleep or hallucinate during hypnagogic and hypnopompic states. On the other hand, if one produces or if one experiences unusually intense internally generated activity that overcomes the processing of reality-based feedback information from the outside world, one also is likely to experience hallucinations. REM sleep clearly fits this bill, but so do a variety of dreamlike states, psychedelic states, temporal lobe epilepsy, and temporal lobe stimulation. Direct stimulation of thalamocortical pathways may also produce hallucination (Llinas, 2001). Even particularly intense wishfulness, such as during food or water deprivation or during an acute mourning period, can produce visual and olfactory hallucinations (Rees, 1971).

work in comparative anatomy and commitment to Darwinian evolution (Ritvo, 1974), he could not escape, in this area, what was accepted doctrine: a neuroanatomical gulf between man and the animals. Freud apparently knew what was at stake, as he said the following in his Open Letter to Dr. M. Fürst concerning the sexual enlightenment of children: "A priest will never admit that men and animals have the same nature, since he cannot do without the immortality of the soul which he requires as the basis for moral concepts" (Freud, 1907, p. 139). Losing corticocentrism would be another blow to Man's narcissism and the privileged place on earth and in the universe that he seemed to need to feel that God had given him.

Following Freud, however, there may have been other forces at work that have delayed what Solms, with the help of affective neuroscience, has recognized. Recall that until the establishment of neuropsychoanalysis there had been, within psychoanalysis, a steady retreat from considering psychoanalysis a natural science. Metapsychology and references to drives were in particular under constant attack. George Klein, Merton Gill, Robert Holt, and Roy Shafer, among others, called for a complete rejection of metapsychology (Ellman, 2010). The biological roots of psychoanalysis were viewed as an embarrassment and as antithetical to the proper, more "humanistic" view of human nature. This movement away from foundational psychoanalytic discoveries has only accelerated in the era of attachment theory, even though neuroscience has so much to offer to help support and refine its theories.

In a recent article by Kaveh Zamanian, "Attachment Theory as Defense: What Happened to Infantile Sexuality?" (2011), he describes the attention paid to "infantile sexuality with its emphasis on the body as the earliest means of emotional regulation and self-experience" as declining and endangered. He references the plummeting interest in psychosexuality in the literature as documented by Fonagy (2008), who says that the resistance to psychosexuality is undiminished and represents an inexcusable collusive negation. Furthermore, a study of psychodynamic therapists by Shalev and Yerushalmi (2009) has shown a shocking discomfort with and ignoring of sexuality. Zamanian (2011) states that there is a "ubiquity of such evidence that marginalization of psychosexuality is a conceptual error with severe technical implications" (p. 39). Apropos this is Alan Schore's comment: "Recent psychobiological and neurobiological studies thus strongly indicate that *the concept of drive, devalued over the last twenty years, must be reintroduced as a central construct of psychoanalytic theory*" (1997, p. 827; emphasis in original).

Correcting the corticocentric bias requires a return to Freud's earliest discoveries about the centrality of bodily experiences and thus to his "rigorous and steadfast attempt to ground mental existence in organismic life" (Opatow, 1999, p. 102). Ironically, those foundational efforts were made at a time when Freud felt it necessary, at least consciously, to disassociate himself from his neurological training. Yet now, when matters of the brain *are* under discussion, psychoanalysis may finally be in a position to warrant a permanent seat at the table.

REFERENCES

Carhart-Harris, R., & Friston, K. (2010). The default mode, ego functions and free energy: A neurobiological account of Freudian idea. *Brain,* 133: 1265–1283.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: Putnam.

Ellman, S. J. (2010). *When Theories Touch: A Historical and Theoretical Integration of Psychoanalytic Thought*. London: Karnac.

Fonagy, P. (2008). A developmental theory of sexual development. *Journal of the American Psychoanalytic Association*, 56: 1–36.

Freud, S. (1891) (with Breuer, J.). *On Aphasia*. New York: International Universities Press, 1954.

Freud, S. (1895). A project for a scientific psychology. *Standard Edition*, 1: 295–387.

Freud, S. (1896). *Extracts from the Fliess Papers*. *Standard Edition,* 1.

Freud, S. (1900). *The Interpretation of Dreams. Standard Edition*, 4/5.

Freud, S. (1907). The sexual enlightenment of children. *Standard Edition,* 9: 129–140.

Freud, S. (1930). *Civilization and Its Discontents. Standard Edition*, 21: 57–146.

Friston, K. (2010). The free-energy principle: A unified brain theory. *Nature Reviews. Neuroscience*, 11: 127–138.

Hobson, A. J. (2009). REM sleep and dreaming: Towards a theory of protoconsciousness. *Nature Reviews. Neuroscience*, 10: 803–813.

Llinas, R., (2001). *I of the Vortex: From Neurons to Self.* Cambridge, MA: MIT Press.

Newberg, A. B. (1994). Changes in the central nervous system and their clinical correlates during long-term space flight. *Aviation, Space, and Environmental Medicine,* 65: 562–572.

Opatow, B. (1999). Affect and the integration of mind and body. *Neuropsychoanalysis,* 1: 97–110.

Rees, W. D. (1971). The hallucinations of widowhood. *British Medical Journal*, 4: 37–41.

Ritvo, L. (1974). The impact of Darwin on Freud. *Psychoanalytic Quarterly*, 43 (2): 177–192.

Schore, A. (1997). A century after Freud's *Project*: Is rap-

prochement between psychoanalysis and neurobiology at hand? *Journal of the American Psychoanalytic Association,* 45: 807–840.
Shalev, O., & Yerushalmi, H. (2009). Status of sexuality in contemporary psychoanalytic psychotherapy as reported by therapists. *Psychoanalytic Psychology*, 26: 343–361.
Solms, M., & Saling, M. (1986). On psychoanalysis and neuroscience: Freud's attitude to the localizationist tradition. *International Journal of Psychoanalysis*, 67: 397–416.
Strachey, J. (1930). Editor's introduction. In: S. Freud, *The Interpretation of Dreams. Standard Edition*, 4: xi–xx.
Zamanian, K. (2011). Attachment theory as defense: What happened to infantile sexuality? *Psychoanalytic Psychology*, 28 (1): 33–47.

The "Conscious Id": A Game Changer with Lots of Challenges

Commentary by Larry Kunstadt (New York)

By integrating contemporary neurobiology with early Freud, Mark Solms opens up the possibility of a major restudying of metapsychology, a project that must continue after his article. My commentary is written not as a critique but as an ally seeking to help Solms fine-tune his article in order to preempt critiques from likely quarters. Four areas are addressed: general issues of approach and clarity; neurobiology; theory/metapsychology; and political issues within the realm of neuropsychoanalysis.

Keywords: affective and cognitive conscious; cortex; clarity of writing, political context; upper brainstem/limbic system neurobiology

The Target Article, "The Conscious Id," is a blockbuster elaboration of basic psychoanalytic theory. The main thrust of this commentary is to suggest some fine-tuning of Mark Solms's proposals in order to preempt critiques from likely quarters. It would benefit greatly from more examples, more definitions, clarifications of terms and concepts, more statements of what experiments would support or refute the claims, and more explication of what this all means both clinically and metapsychologically.

General comments

1. There is some looseness of terms. For example, the fact that upper-brainstem structures are "affective" does not necessarily mean that they are "id." Early on in the article, Solms makes clear why he believes the id should be considered to be located in the upper brainstem (and limbic system), equating it with the homeostatic regulation of the internal body, and he makes clear the role of affect in such regulation. However, thereafter he speaks loosely of "id" and "affect" as if they were interchangeable terms.
2. Freud always said that all affects are conscious. Solms wrote a paper on this claim (Solms & Nersessian, 1999) and repeats this claim in his article. So, it is the neural loci of the two types of consciousness (affective and ideational) that he is revising and elaborating in this article, not the fundamental metapsychology.

   Analysts *today* commonly talk about unconscious affects, but Freud knew that that was an oxymoron. It was just a shorthand way to describe clinical findings, as he mentions in his metapsychological paper, "The Unconscious" (1915, Section III).
3. Psychoanalytic metapsychology was Freud's attempt to explain what he saw clinically. It was not designed to be a general psychology, even though it borrowed from the science and psychology of his time. Solms makes clear at the end of his article that neurobiology cannot prove or disprove clinical data, and vice versa. So in a sense his reliance on neuroscientific data to derive a more accurate metapsychology is confusing domains. I think this is the part of his article that is hardest to get one's arms around because it assumes we have solved the mind/body problem. The distinction between neurobiology and metapsychology needs to be made more distinctive.
4. As Solms knows, instinct is invisible. What counts to the analyst is what is visible—the linguistically represented wish, an id derivative, not the id itself. Analytic metapsychology is first and foremost a theory of thinking, not of instinct, not of

Larry Kunstadt: Institute for Psychoanalytic Education, affiliated with the NYU School of Medicine, New York, U.S.A.

affect, and not of behavior. The currency of the psychoanalytic mind is the wish. The wish that the analyst sees is always attached to a memory. I may have missed it, but it does not appear clear in this article whether Solms considers the id to support an empty ideational consciousness or also its contents.

5. There are two areas we would have all profited from if Freud had addressed them. One is a taxonomy of consciousness. The other is a taxonomy of affects. But he did not propose either. Metapsychology is largely about unconscious meaning, ideation, its transformations and output. (It is based on Freud's modeling of the process of thinking on Wernicke's reflex arc, in his monograph *On Aphasia* (1891), where thinking and some affects are conceptualized as outputs of memory.) If affective consciousness supports ideational consciousness without modulating its contents, then Solms's revisions will have little pragmatic interest to clinicians.

Neuroscientific comments

1. A main contention of the article is that the upper brainstem is affective, part of the id, and that consciousness is preserved in people with lesions above the upper brainstem but not in people with upper-brainstem lesions; therefore, the id generates consciousness. Higher levels may generate specific qualities of consciousness but not consciousness itself.

   These claims are going to provoke enormous controversy from neuroscientists. For a start, the James–Lange theory still has many proponents. They will say that what Solms is describing in cortical lesions is that the upper brainstem stimulates the autonomic nervous system but the cortex can no longer read its signals, so it cannot detect or generate the proper affect. Furthermore, if the upper brainstem itself is lesioned, it cannot turn on the autonomic nervous system, so there is nothing for the cortex to "read." Therefore, it is still the cortex that generates consciousness.

2. As Solms knows, until recently the belief was that a single neural circuit, such as the Papez circuit, mediates affect. Jaak Panksepp, Joseph LeDoux, and other contemporaries have demonstrated that there are different circuits for different affects. I am not sure if it is known yet whether these different affective circuits share a single "consciousness" circuit or whether there are several such circuits. Solms's one-size-fits-all theory is open to criticism that there may be several consciousness networks.

   Unconscious guilt and anxiety play a special role in analysis. I do not think anybody understands their neural basis, but I am willing to bet it will be different from that of the "regular" affects such as love, hate, anger, fear, etc., all of which have several component affects and underlying neural systems. My guess is they will be much more closely related to the systems mediating cognition. Solms may want to distinguish between different affects, especially if he considers them all to be aspects of the id. It is not clear in Solms's conceptualization whether different affects support different types of consciousness.

3. One has to be very careful interpreting extirpation studies, since the effect you get when you remove a structure does not mean that the function was *in* that structure. Solms knows this and has said so many times. So when he states that if the periaqueductal grey (PAG) is lesioned the lights go out, that does not mean that consciousness resides in the PAG. This is a type of localizationist reasoning that Freud argued strenuously against in *On Aphasia* (1891). This reasoning applies to what Solms says about Bud Craig's work and decorticate babies, too.

4. I do not know if Solms is familiar with the article in *Science* by Masataka and colleagues (2011) on vision. In contrast to what everyone has accepted for decades, the authors claim that V1 does *not* mediate visual consciousness. They claim that visual consciousness occurs downstream (V2, V3), and that V1 mediates attention. If this turns out to be true, it opens up the question of whether any of the primary sensory areas mediate the sensations they are classically associated with, or whether raw sensation is a higher derivative integrative function. If so, to support Solms's theory one would need to look again at where the "id" pathways terminate—in primary sensory or association areas?

5. As I am sure Solms also knows, the traditional Jacksonian hierarchical organization used by Freud is only one way of looking at what neural anatomy tells us about neural function.

6. There is also the question of the temporal aspect of consciousness. The importance of limbic loops is, presumably, that they sustain affect over time. I am not familiar with the psychological literature on ideation, but the question would be: "How thick is a human thought?" If affect generates consciousness, then the temporal duration of an affect should have some relation to the duration of a thought. (Perhaps

something such as one affect cycle supports the duration of some number of thoughts.) This is one reason I like Freud's idea that memory forms the basis of all motor output, including consciousness: it eliminates the problem of temporality.

7. How would Solms integrate his new concept of consciousness with his previous work on dream consciousness?

Theoretical comments

1. The main claim Solms makes is that the id generates ego consciousness and that the ego is unconscious. These are really two independent ideas and are certainly separable empirically. The former is the more interesting and easier to accept. I am not even sure I understand what it would mean for the ego to be completely unconscious (everybody accepts that it has an unconscious part). If Solms expects analytic theorists to change this basic aspect of metapsychology, he is going to have to work it out in a lot more detail and not rely on neuroscience to do any more than propel them to think about what it really means metapsychologically. It is ultimately in the clinical arena that the viability of this twofold claim will have to be assessed.
2. Is Solms saying that consciousness, although a derivative of the id, is affectless, just a blank slate waiting for thought to be written on it, or does he accept that the affect can modulate consciousness? Affect certainly modulates the content of consciousness. Maybe we need to more clearly distinguish affective consciousness from ideational consciousness. There is also a certain circularity in the argument—if affects are conscious by their very nature, then how can they support a blank consciousness for thought? And where does the consciousness of affects come from? Perhaps I am being unfair, because we are so far from solving the mind/body problem, but to say they are intrinsic to the brainstem does not really address this issue.
3. The claim that affect creates an empty page of consciousness upon which is written the literature of thought sounds like the concept of the mother's breast as a "dream screen" (Lewin, 1946, 1953) upon which content is projected and the ontogenetically related concept of hallucinatory wish-fulfillment, which Solms discusses. In Ralph Ellison's novel *Invisible Man* (1947) he captures the dynamic aspect of blank consciousness as "voices without messages, of newsless winds."

   However, to make it very short, Opatow's (1997) interpretation of Freud is that it is the wish—the thought, the cognitive representative of the bodily state—that causes consciousness, not the affect itself. He also makes a distinction between a passive consciousness and an active consciousness, which creates a gap between internal and external reality. (He has a long and complex argument.) Where in Solms's conception is there a role for the unconscious wish? For unconscious memory?
4. Is Solms suggesting that you need a real life experience, such as the mother as object, to create consciousness? I know this is an area of contention because it is next to impossible to study empirically, but isn't a baby born with consciousness? I would surmise that a baby has sensory consciousness before birth, although maybe not thought. Does Solms think it is the birth experience that stimulates the affect that causes consciousness? I am not sure Freud was right on this issue.
5. The distinction between affect determining or distorting thinking and supporting thinking needs to be fleshed out. If I understand Solms correctly, he is arguing that affect allows ideational consciousness to create the structured space needed for thinking, not that affect supports the meaning of the thinking process, which, in fact, it may distort.
6. Affect that represents the bodily state, which is what Solms is talking about, may be different from affect as a form of memory, which is what gives people their personalities. It is the repetition of fixated or regressed infantile affective memories that is so noticeable in pathology.
7. In Section 3, "The Corticocentric Fallacy," Solms reminds us that Freud said that only a perception can become conscious. Solms, by contrast, seems to argue that consciousness reads motor output. I am not clear what he is getting at here. I think that Freud meant to make this claim only in relation to memory—Freud knew that perception, as Jason Brown (1988, Part II) also says in his model, is a motor act, that perception is really an awakened modulated memory reprojected outward onto the world. This would result in a perception of this motor behavior. I do not think Freud was trying to say anything here about external perception. Is Solms's point that all instincts are unconscious?
8. What bothers me most of all about all the theories of consciousness I have seen is that they do not capture the dynamic aspect of it: that it needs to be continuously generated—it is not just there, the

way an object in the external world is. I think Solms understands this, although it is not emphasized. This is one aspect of consciousness that could also be studied empirically.

Political comments

Some papers are ignored, some get a flurry of attention and then disappear, and some make a lasting impression. Solms's article has the potential to become a foundational paper, studied and debated for decades, which, I assume, is what he wants.

There are several differences between this article and many of those that have disappeared. One is that many innovative papers in psychoanalysis are scarcely intelligible, whereas this one is accessible. The second is that Solms has a preexisting infrastructure of interested people, the neuropsychoanalytic movement. The third is that his article is backed by empirical evidence, which many innovative papers in psychoanalysis are not. The fourth is that Solms is generally recognized as the leader of the movement; so this article is likely to find an existing sympathetic audience.

However, there are some things Solms can do to keep this paper on people's radar screen.

- Solms must give people something to test empirically. He must bring them in. He must give people the chance to spend the next years testing these ideas. He must be concrete. He must say what studies would prove or disprove the assertions.
- Many people today do not pay attention to metapsychology. Solms has to show them why it counts (why theory is important for clinical work). So, he has to show them the clinical implications of his revision. He mentions at the end that people should consider using the concept clinically, but he does not say how. For non-analysts, Solms must show how radically different in some respects this proposal is from mainstream cognitive neuropsychology, so they have something to yell and scream about. Nobody accepts something someone else says; people only learn when they discover it for themselves.
- I would also suggest that Solms should give more neuroscientific examples that affirmatively support his view of the conscious affective id. His argument is based on only two studies with no postmortem data.
- I was very struck by the part of the article that summarizes some of the contradictions in Freudian theory that Solms's proposal would make consistent. However, few people care about such matters; getting rid of Freud's contradictions is on few people's mind these days, although it certainly is a salutary consequence of his proposal.

The Target Article is a real *tour de force*. Neuropsychoanalysis is now about 20 years old. My guess is there are perhaps several hundred people, maybe more, who in some way contribute to the field. Everybody else's work more or less articulates with psychoanalysis, but Mark Solms's and Barry Opatow's articles are the only ones that actually integrate neurobiology and psychoanalysis! It is a privilege to study this article. I look forward to future articles that explore the metapsychological consequences in detail.

REFERENCES

Brown, J., (1988). *The Life of The Mind.* Hillsdale, NJ: Lawrence Erlbaum.

Ellison, R. (1947). *Invisible Man.* New York: Vintage, 1995.

Freud, S. (1891). *On Aphasia.* New York: International Universities Press, 1953.

Freud, S. (1915). The unconscious. *Standard Edition*, 14: 177–179.

Lewin, B. (1946). Sleep, the mouth, and the dream screen. *Psychoanalytic Quarterly*, 15: 419–434.

Lewin, B. (1953). Reconsideration of the dream screen. *Psychoanalytic Quarterly,* 22: 174–199.

Masataka, W., Cheng, K., Murayama, Y., Ueno, K., Asamizuya, T., Tanaka, K., et al. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science,* 334 (6057): 829–831.

Opatow, B. (1997). The real unconscious: Psychoanalysis as a theory of consciousness. *Journal of the American Psychoanalytic Association*, 45: 865–890.

Solms, M., & Nersessian, E. (1999). Freud's theory of affect: Questions for neuroscience. *Neuropsychoanalysis*, 1: 5–14.

What Is Consciousness? A Tridimensional View and Neural Predispositions of Consciousness (NPC)

Commentary by Georg Northoff (Ottawa)

Mark Solms raises the issue of the neuronal and conceptual characterization of consciousness. He focuses very much on stimulus-induced activity in relation to affective and cognitive functions. This, though, implies a content-based view of consciousness that defines consciousness by its contents—that is, affective and cognitive. Beside content, recent discussions often consider the level of consciousness associated with brainstem/midbrain as a second dimension of consciousness. However recent data about the intrinsic activity of the brain suggest the need to include a third dimension—form (or structure or organization)—in the characterization of consciousness. The commentary spells this out, including the implications for the neural correlates of consciousness and Solms's view.

Keywords: consciousness; contents; form/structure; intrinsic activity; neural predispositions

Background: what is consciousness?

Mark Solms's impressive Target Article about the role of the body and its relation to two different forms of consciousness—affective and cognitive—raises many important empirical and conceptual issues. I shall focus here on only two main points. First, the characterization of consciousness by specific contents—for example, affective and cognitive—may need to be complemented by yet another dimension, form or structure, or organization, as one may want to say. Second, this implies that we may need to look for neural predispositions of consciousness (NPC) rather than neural correlates of consciousness (NCC) (Northoff, 2011, 2012a, 2012b, 2012c, 2012d). While Solms extends the reach of the NCC to affective consciousness, he still remains within the domain of contents while not really providing an answer regarding the NPC.

While there have been many suggestions for the different cognitive forms of consciousness, the neuronal mechanisms underlying the subjective and phenomenal-qualitative dimension, and thus phenomenal consciousness, remain unclear. The focus in this commentary is therefore on phenomenal consciousness, which is also implied when using the term "consciousness" by itself. Such phenomenal consciousness can occur in association with a variety of content—cognitive, sensorimotor, and affective—with the latter being pointed out by Solms.

Georg Northoff: University of Ottawa Institute of Mental Health Research, Ottawa, Canada.

Acknowledgments: I am thankful to the CIHR, the ISAN-HDRF, the CIHR-EJLB, and the Michael Smith Foundation for generous financial support.

Content and level of consciousness—a bidimensional view

Consciousness is usually considered in a bidimensional way by content and level. How can we describe content and level of consciousness? Content refers to events, persons, or objects that are associated with consciousness. Much neuroscientific research has focused on the neural mechanisms that allow the transfer of content from the unconscious to consciousness. Several suggestions for this have been made, including cyclic thalamocortical reentrant processing, information integration, global neuronal workspace, prestimulus resting-state activity, low-frequency fluctuations, and neuronal synchronization.

These neuronal mechanisms are assumed to be sufficient to associate content with consciousness. They are thus considered what is typically described as "neural correlates of consciousness" (see Chalmers, 2010; Crick, 1994). Since they refer specifically to the contents of consciousness, one may also speak of content-based NCC. This is also relevant in neurological disorders that most often can be characterized by lesions in specific regions that process specific contents. The visual content-based NCC are, for instance, impaired in patients with selective lesions in V1 or V5.

In addition to content, level of consciousness describes the different degrees or states of consciousness. More specifically, the level of arousal and awakening as indicated by active reaction and behavior toward the environment signifies the level of consciousness. The level of consciousness has been associated with neural activity in the brainstem/midbrain and the brain's global metabolism. Solms now adds another view, in that he also associates specific contents with subcortical regions—namely, affective contents—as this is well in line with Panksepp (2011).

Form of consciousness—a tridimensional view

How are the brain's intrinsic activity and its spatiotemporal continuity related to the contents of consciousness? The brain's intrinsic activity shows a certain temporal and spatial structure. The temporal structure is realized by the high- and low-frequency fluctuations of its activity level, whereas the spatial structure is realized by the functional connectivity between different regions (for details, see Northoff, 2012a, 2012b). The spatiotemporal continuity seems to structure and organize the contents in time and space. The contents are embedded into a spatial and temporal context. More specifically, the discrete points in time and space associated with particular contents are integrated and embedded into a context of ongoing spatiotemporal continuity.

The spatiotemporal continuity may be central in integrating and thus structuring and organizing the contents of consciousness in time and space. As described above, space and time provide a matrix of spatiotemporal continuity. Such a matrix allows us to integrate, structure, and organize the various contents within a spatiotemporal context. One may thus say that the spatiotemporal matrix provides a form—the form of spatiotemporal continuity—for consciousness and its various contents.

What exactly do I mean by "form" of consciousness? The concept of form has been used in philosophy since the time of ancient Greece, where it was distinguished from the contents or the actual material. Nowadays, one may want to describe the concept of form with the terms "structure" and "organization," as they provide some kind of grid or template. The form allows us to "put together" the different contents in and across their different discrete points in time and space. One may thus want to define the form of consciousness as "putting together." Such "putting together" may be specified as the "structuring and organizing of contents in time and space."

How is such form related to the level of consciousness? As discussed above, level of consciousness concerns the state or degree of consciousness as manifest in arousal and awakening. This is to be distinguished from the form that concerns the spatiotemporal organization of consciousness. Hence, form may need to be distinguished from the level of consciousness as a distinct dimension.

Taken together, in addition to the content and level of consciousness, I describe here yet another dimension of consciousness, opting for a tri- rather than bidimensional view. This form of consciousness refers to the structure and organization of consciousness and, more specifically, to how the different contents are "put together" in time and space.

Thereby, the content's discrete points in time and space are seen as integrated and embedded into a spatiotemporal continuity that can be considered as an organizational matrix or structure. Neuronally, such a spatiotemporal continuity may be assumed to be traced back to the intrinsic activity—more specifically, to the intrinsic activity's constitution of spatiotemporal continuity across neural activities at different discrete points in time and space.

Neural predispositions of consciousness

The concept of the neural correlates of consciousness describes the nonnecessary, sufficient neural conditions of consciousness. Thereby, most of the suggestions for the NCC (see above) focused mainly on the content of consciousness when aiming for the neural mechanisms that allow one to distinguish unconscious and conscious content.

The NCC have recently been separated into the distinction between neural prerequisites, neural substrates, and neural consequences of consciousness (see Aru, Bachmann, Singer, & Melloni, 2012; deGraaf, Hsieh, & Sack, 2012). Neural prerequisites describe neural processes that are necessary but not sufficient by themselves for consciousness to occur. As such, neural prerequisites must be distinguished from neural substrates that concern those neural mechanisms that are directly related to the conscious experience itself. Finally, neural consequences describe those neural processes that result directly from the neural activity underlying consciousness.

How does that compare to the suggested central role of the intrinsic activity providing the form of consciousness? One may consider the intrinsic activity and its spatiotemporal activity pattern as a neural prerequisite that is necessary for consciousness to occur. At the same time, while it may not be sufficient by itself, it would fulfill the definition of being a neural prerequisite of consciousness.

However, the brain's intrinsic activity is not only a necessary condition for the manifestation of a particular conscious state but also for consciousness in general. As expressed in philosophical terms, the intrinsic activity is a necessary condition of possible consciousness. As such, it may be regarded as what I described recently as the neural predisposition of consciousness (NPC; Northoff, 2011, 2012a, 2012b, 2012c, 2012d). The concept of the neural predisposition of consciousness describes the conditions that make or design those brain states favorable (rather than nonfavorable) to constitute consciousness (rather than nonconscious). The NPC can thus be said to underlie the brain's tendency to create or constitute consciousness.

Let us draw an analogy to the heart. Just as the heart's intrinsic muscle structure predisposes it to pump (rather than to not pump), the brain's intrinsic activity predisposes the brain to constitute consciousness (rather than nonconscious). The analogy goes even further. In the case of the heart, additional features are needed to make the heart pump blood—namely, electrical discharges to activate contractions of the muscles, which then allows for pumping. Hence, by itself, the muscle structure cannot do anything.

This may be true, too, in the case of the brain and its intrinsic activity. In addition to the intrinsic activity and its spatiotemporal pattern, additional neural processes such as the neural prerequisites and the neural substrates are needed to instantiate consciousness. However, the neural prerequisites and neural substrates themselves would not be able to instantiate consciousness without the intrinsic activity and its spatiotemporal pattern. The intrinsic activity and its spatiotemporal pattern do, therefore, predispose possible consciousness thus being an NPC as distinguished from the NCC and their threefold distinction into neural prerequisites, neural substrates, and neural consequences.

Form as the neural predisposition of consciousness

How do the NPC relate to the suggested threefold distinction between content, form, and level of consciousness? The NCC have focused much on content of consciousness. The extension of the NCC into neural prerequisites introduces the distinction between content-variant and content-invariant processes. The intrinsic activity and its spatiotemporal pattern is certainly content-invariant, which is supported by the fact that it occurs considerably prior to any specific content entering consciousness. The NPC thus does not concern the constitution of content and must therefore be distinguished from all related neural processes.

How about the NPC and the level of consciousness? As discussed above, the intrinsic activity and its spatiotemporal pattern are dependent on the energy and glucose metabolism associated with the level of consciousness. However, imagine the case where there would be complete and sufficient energy and metabolic supply but without the constitution of a spatiotemporal pattern in the brain's intrinsic activity. In that case, consciousness would still remain absent and be principally impossible. This strongly suggests that the NPC cannot be determined by the level of consciousness.

Taken together, the NPC are strongly aligned to the form of consciousness, the intrinsic activity, and the spatiotemporal pattern. Without the intrinsic activity's spatiotemporal pattern providing the form of consciousness, consciousness would be altogether impossible. There would only be unconscious, and such unconscious would no longer in principle have consciousness. In more technical terms, the NPC account for the form and its underlying neural mechanisms as the necessary condition of the principal possibility of consciousness—for example, possible rather than actual consciousness.

Conclusion

How does all the above stand in relation to Solms and his "godfather" Freud? Solms enlarges the reach of the content-based NCC by making a case for affective consciousness. That, though, addresses only the question of the kind of contents of consciousness and their relation to neural activity—for example, the NCC. He thereby targets the distinction between conscious and unconscious contents specifically with regard to affective contents.

However, he leaves open why there is consciousness at all rather than nonconscious, a question that is described as the hard problem in the current philosophy of mind. For that, as I argue, we need to include a third dimension of consciousness: the form or structure or organization of the brain's intrinsic activity. That form, applied by the brain's intrinsic activity to any of its neural processing of extrinsic stimuli and their associated contents, may predispose the brain to associate the various contents—sensorimotor, affective, and cognitive—with consciousness.

This would explain why all functions and their related contents—sensorimotor, affective, and cognitive—can be associated with consciousness. Hence, it is the form or structure of our brain's intrinsic activity rather than its contents and their related function (sensorimotor, affective, cognitive) that we need to understand in order to unravel why there is consciousness rather than nonconscious.

And that means, ultimately, to do nothing else than to apply and transfer Freud's concept of structure from his context of the psychic apparatus to the brain and its intrinsic activity. In the same way that Freud understood the structure of the ego as a psychological predisposition for the kind of contents we can consciously experience, the brain's intrinsic activity may provide the structure for the neural processing of extrinsic contents and their possible association with consciousness.

REFERENCES

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlate of consciousness. *Neuroscience and Biobehavioral Reviews,* 36: 737–746.

Chalmers, D. J. (2010). *The Character of Consciousness.* Oxford: Oxford University Press.

Crick, F. (1994). *The Astonishing Hypothesis.* New York: Simon & Schuster.

deGraaf, T. A., Hsieh, P. J., & Sack, A. T. (2012). The "correlates" in neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews*, 36: 191–197.

Northoff, G. (2011). *Neuropsychoanalysis in Practice: Brain, Self, and Objects.* Oxford: Oxford University Press.

Northoff, G. (2012a). *Unlocking the Brain, Vol. I: Coding*. Oxford: Oxford University Press.

Northoff, G. (2012b). *Unlocking the Brain, Vol. II: Consciousness*. Oxford: Oxford University Press.

Northoff, G. (2012c). Immanuel Kant's mind and the brain's resting state. *Trends in Cognitive Sciences*, 16 (7): 356–359.

Northoff, G. (2012d). *Das undisziplinierte Gehirn. Was nun, Herr Kant?* [The undisciplined brain: What now Mister Kant?]. Munich: Random House.

Panksepp, J. (2011). Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloS ONE,* 6 (9), e21236. doi:10.1371/journal.pone.0021236

Toward an Understanding of the Constitution of Consciousness Through the Laws of Affect

Commentary by Jaak Panksepp (Pullman, WA)

Abundant evidence, often ignored in discussions of cognitive consciousness, is that raw affective experiences arise from diverse subcortical emotional, motivational (body homeostatic), and primal sensory systems (e.g., taste and smell). These primary-process affective systems that generate diverse types of valenced feelings may not be well described by the common synonym for consciousness called "awareness," which is a better descriptor for higher forms of consciousness—namely, knowing that you experience. What a subcortical affective consciousness offers for our understanding of the mind is a primal form of phenomenal experience, which may be the foundation for the rest of the mind. From a neuropsychoanalytic perspective, Mark Solms advances a thesis for understanding the ancestral sources of mind that is consistent with data on cross-species emotional systems of all mammalian brains.

Keywords: affects; cognitive consciousness; emotions; hard problem; phenomenal consciousness; SEEKING

Little heed has been paid in consciousness studies to the likelihood that the primal affective *qualities* of mind serve as a foundation for the rest of the mental apparatus. In "The Conscious Id," Mark Solms shares provocative visions of how the ancient (subcortical) evolutionary strata of affective consciousness—endogenous qualities of mind—percolate up through more recent developments in the neural fabric to control the construction of higher cognitive minds. This remains a novel vision in consciousness studies. Solms weaves his tapestry of possibilities, radical for both neuroscience and psychoanalysis, with exceptional clarity and devotion to facts rather than the usual hand-me-down culturally condoned opinions, and he shares these novel ideas with the humility befitting such radical transformations in traditional understanding. But even with such conceptual advances, we still need to reverse the poverty of knowledge concerning the actual *constitution* of consciousness.

What Solms shares may not sound right to many who still believe the top of our brains generates consciousness. However, the need for such a rethinking has been growing, almost unrecognized, for decades. For scores of years, my efforts have been devoted empirically to clarifying the nature of affective consciousness, and Solms's article exquisitely shares how the resulting transformations of our understanding of even our higher intrapsychic processes might be brought

Jaak Panksepp: College of Veterinary Medicine, Washington State University, Pullman, WA, U.S.A.

into line with the current evidence. Perhaps this can help us develop more effective psychotherapies that recognize that what may be cognitively unconscious may not be affectively so (see Shevrin, Panksepp, Brakel, & Snodgrass, 2012)—allowing implicit cognitive memories to wreak havoc in emotional lives when the precipitating causes have been long forgotten or repressed (no longer retrieved for long spans of time).

We await to see how this radical transformation of psychoneurophenomenology evolves, through many scientific and cultural resistances—hopefully not as robust as those I experienced while restoring affective consciousness, based on convergent evidence, to the other animals that were left outside the "circle of affect" by behavioral scientists a century ago. My radical departure was pursued largely to achieve a better understanding of the constitution of our human emotional feelings. I sought to counter the existing madness engendered by behaviorism by emphasizing one robust finding of modern neuroscience: Wherever in the brain one evokes coherent emotional-instinctual behavior patterns with deep brain stimulation (DBS), those shifting states serve as rewards and punishments in simple approach and escape tasks. This gives us evidence for affective consciousness, permitting us to envision new "Laws of Affect," where the circuits of emotional feelings controlled learning, but there is still a long way to go in order to understand what kind of neural activities actually instantiate the affective experiences (Panksepp, 2000). But the subcortical loci of control have been identified (Panksepp, 2005a, 2005b).

Why are there still such vast cultural resistances to such straightforward evidence that Solms summarizes, especially among scientists who should be playing by the "rules-of-evidence"—the most important being the "weight of affirmed predictions"? Perhaps, because as our higher cognitive mind become ever more conversant with complex ideas, as opposed to just the affective-perceptual coloring of our external worlds that anencephalic children possess, there emerges a tendency to reverse foundational and derivative issues in psychology and consciousness studies. For instance, by adhering to the James–Lange tradition still favored in psychology (although it is not consistent with most neuroscience evidence), most psychologists still believe that our affective feelings—the unconditioned passions of our minds—reflect cognitive-type "readouts" of unconscious autonomic arousals within our higher mental apparatus, rather than being intrinsic dynamics of our subcortical neural networks.

Thus, within the Western intellectual tradition, which bled readily into Freudian thought, consciousness was envisioned as an emergent of higher brain functions, rather than a fundamental property of lower brain regions that elaborate affectively experienced states of being (Panksepp, 1998a, 1998b, 2007; Panksepp & Biven, 2012; Solms & Panksepp, 2012). Still, the subcortical origin of primary-process affects leaves open the possibility that cognitive consciousness arose directly from the neocortex being in touch with our subcortical sensory-perceptual proto-experimental capacities much more than our affective ones (Merker, 2007). How might that be negated? Until such issues can be experimentally evaluated, the origins of cognitive consciousness will remain murkier than the origins of affects.

In any event, surely we must first experience ourselves as living creatures before we can envision ourselves as part of the greater world, but the neural transitions from intrinsic affective potentials to fuller cognitive abilities remain to be mapped. While anencephalic children and animals certainly provide compelling evidence for how much mind exists in subcortical domains, this does not unambiguously indicate that perceptual qualia emerged from affect, albeit I remain fond of that idea (Panksepp, 1998b, 2007). One implication of Solms's analysis that affect may provide the neural grounding for cognitions is that the transition from an egocentric affective perspective to a fully human world-centric cognitive perspective is never complete. We all remain affectively self-referential to some extent. What are the neuropsychoanalytic implications of this surprising turn of event, if true?

How might we construct detailed cognitive visions of ourselves and our worlds, both external and internal, from our diverse affective experiences of the world based largely on sensory inputs, which are presumably transformed into cognitive representations, through the power of affect? Of course, no one knows, but where might we go experimentally or clinically with this central idea? Do we need to commit to the evolutionary view that what came first in BrainMind evolution sustains priority throughout our lifespans? If we consider the metaphor of well-rooted plants, is it the case that the rich potential for cognitive foliage already resides within the neonatal seed sprout, which can never mature without the mental nourishment arising from subterranean roots (a metaphor for the subcortical affective mind)? If that is the case, the deep and empirically unanswered questions, perhaps ready for harvesting (and surely necessary for future progress), might be: Which of the emotional primes that generate the evolved endogenously affective experiential potentials of our brains are most critically important for our cognitive development? If forced to make such a choice, I would surely select the SEEKING system as the one that would have primacy, as described in the last issue of this journal (Wright & Panksepp, 2012). I would predict that if this enthusiasm- and foraging-promoting brain system were severely damaged, higher mental life would not emerge in any infant.

Surprisingly, such bottom-up visions of the evolution of mind remain minority positions within consciousness studies, where more empirical effort is devoted to the study of the neural *correlates* of perceptual consciousness (Crick & Koch, 1998) rather than the neural *constitution* of consciousness (for a discussion of this distinction, see Miller, 2007; Panksepp, 2013). Solms's concern, as mine, is more with the latter than the former, which to my way of thinking is essential for any progress toward a "causal" neuroscientific understanding of consciousness. Of course, the key issue is how we might cash out this vision with future neuroscientific research on this topic of ultimate concern. Our position, perhaps the optimal entry-point for scientifically illuminating this "hard-problem consciousness," is that the key to neuroscientific progress may be through the study of the intrinsic emotional processes of mammalian brains.

In short, our perspective is that without the capacity to feel key survival issues—reflecting fundamental affective experiences of the BrainMind—other experiences could not emerge in animal or human brains. If so, without a focus on the nature of primal affective consciousness, we may never penetrate the "hard-problem" of consciousness studies: understanding the phenomenal qualia from which higher forms of cognitive consciousness are constituted. I will now briefly focus on this issue, in the hope of pushing forward the novel agenda that Solms crystallizes in his Target Article. Although the most difficult aspect of this problem may be to scientifically characterize how cognitive consciousness emerges from those lower brain functions, before that can be achieved we do still need a much better understanding of affective consciousness—the neural nature of affective qualia—a "hard-problem" that, I believe, is currently solvable. Without more foundational knowledge about affective experiences, I suspect we simply cannot know how the lower subcortical substrates of mind actually permit higher perceptual and cognitive functions to emerge. Without such "constitutional" knowledge, we will be left in the traditional realm of speculation rather than scientific knowledge. So how do we obtain it? For me, the key is cross-mammalian brain research.

Animal models can dramatically clarify the fundamental nature of the raw emotional feelings of human beings. Such preclinical research provides direct access to the affective circuits that are important in understanding human psychiatric disorders (Panksepp, 2006, 2012). The epistemology has at least half a dozen steps:

1. Identification of emotion-mediating brain networks, which, across diverse species of mammals, generate positive and negative hedonic effects (as with DBS, whether electrical or optogentic; for an overview of optogentics, see *Biological Psychiatry,* 2012).
2. Identification of the major additional brain regions that are most strongly impacted by such brain arousals (e.g., cfos histochemistry for short-term arousals and cytochrome oxidase studies for long-term influences: see Harro, Kanarik, Matrov, & Panksepp, 2011).
3. Identification of the neural connectivities and neurochemistries of these systems.
4. Monitoring neurodynamic correlates from these brain regions (as with quantitative EEG, and small-animal fMRI and PET procedures) as a function of intensity and duration parameters of relevant brain stimulations and behavioral/affective changes.
5. Identification of key chemistries that modify—that is, attenuate and intensify—the above behavioral/affective effects in animals as well as in humans (to the extent possible, in elective informed-consent neuropsycho-surgery to implant cerebral pacemakers) (see Coenen, Schlaepfer, Maedler, & Panksepp, 2011).
6. Providing additional testable/falsifiable predictions, especially for humans, from the above cross-species neuroanatomical, neurochemical; and neuropharmacological findings.

With such a foundation, progress toward cracking the really hard problem of how affective consciousness gets translated into the massive, but initially unconscious, neural networks of the neocortex become manageable. This cross-species affective neuroscience approach to understanding primal affective consciousness has barely begun, partly because "affect" has the supposedly "bad" connotation of consciousness existing in animals, especially among many scientists who have the best tools to make progress on the neuroscientific problem that is most fascinating to many, while most animal behaviorists choose to remain in denial about this issue, especially those who pursue fear conditioning—which may cause them practical problems.

In sum, endogenous internal affective states, from hunger and thirst to diverse emotions, arise from ancient, evolutionarily-dedicated medial brainstem and diencephalic networks that have intrinsic mental abilities (Denton, 2006; Panksepp, 1998a, 1998b, as updated in Panksepp & Biven, 2012). By contrast, the more "objectified" aspects of cognitive consciousness arise from more spacious, but initially affectively "empty" neocortical networks, which developmentally allow higher vertebrates to become cognitive creatures conversant with the complexities of the world. They apparently achieve a higher mind through (1) experience-based acquisition of refined sensory perceptions, (2) mental time-travel through individualized memories, (3) permitting creativity, thoughtfulness, cognitive intelligence, executive control of behavior, and (4) an "awareness" that promotes explicit decision-making, not to mention love, hate, and empathy. If true, we must all begin to ask: How could all that cognition we admire emerge from affective foundations? Solms points us toward an intriguing possibility entertained by Friston and colleagues, but there are surely other options not yet considered. Might Solms provide further guidance on additional explicit experimental paths, especially neuro- and psychodynamic predictions, that this novel vision might inspire?

REFERENCES

Biological Psychiatry (2012). A brief summary of the articles appearing in this issue of *Biological Psychiatry*. *Biological Psychiatry,* 71 (12): 1027–1089.

Coenen, V. A., Schlaepfer, T. E., Maedler, B., & Panksepp, J. (2011). Cross-species affective functions of the medial forebrain bundle: Implications for the treatment of affective pain and depression in humans. *Neuroscience & Biobehavioral Reviews*, 35: 1971–1981.

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex,* 8: 97–107.

Denton, D. (2006). *The Primordial Emotions: The Dawning of Consciousness.* New York: Oxford University Press.

Harro, J., Kanarik, M., Matrov, D., & Panksepp, J. (2011). Mapping patterns of depression-related brain regions with cytochrome oxidase histochemistry: Relevance of animal affective systems to human disorders, with a focus on resilience to adverse events. *Neuroscience & Biobehavioral Reviews*, 35: 1876–1889.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30: 63–81.

Miller, S. M. (2007). On the correlation/constitution distinction problem (and other hard problems) in the scientific study of consciousness. *Acta Neuropsychiatrica*, 19: 159–176.

Panksepp, J. (1998a). *Affective Neuroscience*. New York: Oxford University Press.

Panksepp, J. (1998b). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the SELF. *Journal of Consciousness Studies*, 5: 566–582.

Panksepp, J. (2000). The neurodynamics of emotions: An evolutionary-neuro- developmental view. In: *Emotion, Self-Organization, and Development*, ed. M. D. Lewis & I. Granic. New York: Cambridge University Press, pp. 236–264.

Panksepp, J. (2005a). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition,* 14: 30–80.

Panksepp, J. (2005b). On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies,* 12: 158–184.

Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Progress in Neuro-Psychopharmacology & Biological Psychiatry,* 30: 774–784.

Panksepp, J. (2007). Affective consciousness. In: *The Blackwell Companion to Consciousness,* ed. M. Velmans & S. Schneider. Malden, MA: Blackwell, pp. 114–129.

Panksepp, J. (2012). The vicissitudes of preclinical psychiatric research: Justified abandonment by big pharma? *Future Neurology*, 7: 1–3.

Panksepp, J. (2013). The neuroevolutionary sources of mind: The ancestral constitution of affective consciousness and the Core-SELF. In: *The Constitution of Consciousness, Vol. 2: Toward a Science and Theory of Phenomenal Consciousness,* ed. S. Miller. Amsterdam: John Benjamins.

Panksepp, J., & Biven, L (2012). *The Archaeology of Mind.* New York: W. W. Norton.

Shevrin,, H., Panksepp, J., Brakel, L. L. A. W., & Snodgrass, M. (2012). Subliminal affect valence words change conscious mood potency but not valence: Is this evidence for unconscious valence affect? *Brain Science,* 2: 504–522.

Solms, M., & Panksepp, J. (2012). The "Id" knows more than the "Ego" admits: Neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science,* 2: 147–175.

Wright, J. S., & Panksepp, J. (2012). An evolutionary framework to understand foraging, wanting, and desire: The neuropsychology of the SEEKING System. *Neuropsychoanalysis*, 14: 5–39.

Self-Specificity of the External Body

Commentary by Manos Tsakiris (London)

Mark Solms suggests that the internal body, equated with the id and represented at the brainstem structures, gives us the self-as-subject consciousness, while the external body, equated with the ego, is represented as an object, analogous to any other object in the world. Is memory space the sole, or at least the most important, contribution that the ego can make to the id? I would like to argue that it is not. Even though the basis of phenomenal consciousness—the "being-me" state—might be given by the brainstem consciousness, the most important function of the ego is precisely that it can represent *my* body as an object and *identify* it with the internal body. Thus, both bodies need to be represented as self-specific, and inevitably this will require the contribution of a cortical network. Both the ego and the id, in Solms's terms, co-constitute self-specificity.

Keywords: body-representations; interoception; multisensory state; prediction error; self-specificity

Mark Solms puts forward an intriguing hypothesis that can have far-reaching consequences for psychoanalysis and neuroscience alike. Contrary to what the Freudian metapsychology would suggest, Solms is proposing that the id is the fount of consciousness, while the ego is unconscious in itself. To support this hypothesis, Solms provides us with a new reading of the current knowledge of embodiment as studied in neuropsychology to suggest that the functional integrity of the cortex that represents the external body, the one that was used by Freud in describing how "the Ego is first and foremost a bodily Ego" (Freud, 1923, p. 26), depends on brainstem structures that represent the internal body. In short, without the arousal system that is used to neurally represent the internal body, we would not have consciousness of the body as an external object.

How many bodies in the brain?

From the outset, Solms correctly distinguishes between body representations that relate to the external body, or the body as perceived exteroceptively, and body representations that relate to the body's internal milieu, or the body as represented interoceptively. Interestingly, Solms notes, there is nothing special about the external body insofar as its self-specificity is concerned: the external body that happens to be mine is cortically represented in an equivalent way as other bodies/objects that exist in the world. The proposed functional equivalence between this body that happens to be mine and that body that I see passing by is based on the discovery of mirror neurons and the subsequent conceptualization of a "mirroring system" of bodily movements, as well as of sensory bodily states as more recent studies have shown. On the other hand, it seems that the internal body is self-specific, not only because of the private interior space *within* which it exists, but also because its consciousness "consists in *states* rather than *objects* of consciousness," as Solms notes. Moreover, the consciousness of the internal states is inherently affective, and not perceptual, as consciousness of objects is considered to be. For scholars of consciousness studies, Solms's use of "consciousness" will inevitably raise a series of challenging questions about what is actually meant by consciousness. My understanding is that the brainstem consciousness associated with the id is the phenomenal affective consciousness and that the internal body is the subject of perception, the condition that makes consciousness of the external world possible. In later parts of his article, Solms explains that the external body, an exemplar of perceptual consciousness equated with ego, contributes the much-needed memory space required for stabilizing perceptual objects.

The ego and the question of self-specificity

Is memory space the sole, or at least the most important, contribution that the ego can make to the id? I would like to argue that it is not. Even though the basis of phenomenal consciousness, the "being-me" state, is given by the brainstem consciousness, the most important function of the ego is precisely that it can represent *my* body as an object and *identify* it with the internal body. On the topic, Solms writes: "The subject of consciousness identifies itself with its external body (object-presentation) in much the same way as a child projects itself into an animated figure it controls in a

Manos Tsakiris: Laboratory of Action & Body, Department of Psychology, Royal Holloway, University of London, London, U.K.

television game. The representation is rapidly invested with a sense of self, although it is not really the self." True, the cortex can play tricks and allow identification with all sorts of funny objects, such as rubber hands and mannequin bodies or even unfamiliar faces, as my own studies and those of Henrik Ehrsson and Olaf Blanke suggest (Blanke & Metzinger, 2009; Petkova & Ehrsson, 2008; Tajadura-Jiménez, Grehl, & Tsakiris, 2012; Tsakiris, 2010). Similarly, in the video-game paradigm evoked above, the child might identify with an avatar, which is "not really the self." The reason why the avatar is not the real self is because the very moment the child switches off the screen, the avatar's body disappears, rendering the identification obsolete. However, there is something that is my *real* external body, a *sensitive* object of a certain volume and form that exists in physical space and that, interestingly enough, never leaves *me*. As William James noted, contrary to the perception of an object, which can be perceived from different perspectives or even cease to be perceived, we experience "the feeling of the same old body always there" (James, 1890, p. 235). Echoing James, Merleau-Ponty (1962) wrote that "the permanence of my own body is entirely different in kind. . . . Its permanence is not a permanence in the world, but a permanence on my part" (p. 90). Thus, my external body might be an object, but still it is an object unlike any other, at least "on my part." In addition, the very fact of embodiment—and here I refer to the external body—suggests that more often than not we do have one permanent body of evolutionarily prescribed characteristics. Inevitably, this human body that has evolved through phylogeny comes with some innate priors that must somehow represent its specificity: this one body for this brain.

What I would like to suggest is that, to a certain extent, self-specificity—and not the illusion of self-specificity as described in Solms's avatar paradigm—needs also to arise at the cortical level. This seems necessary for the id, but also for the relation of the ego with the external world. Before addressing this dual necessity, two clarifying remarks are needed. First, several studies have used bodily illusions such as the rubber-hand illusion and the full-body illusion to highlight the processes that underpin consciousness of this external body as mine. The success of eliciting these bodily illusions should not be considered as prima-facie evidence that the external body is an illusion. Instead, such illusions can be thought of as a model instance of the real experience of embodiment, in the same way that somatoparaphrenic symptoms in brain-damaged patients serve the same purpose. Second, notwithstanding the success of inducing illusory body-ownership over foreign external objects, all studies converge on the hypothesis that what drives our consciousness of our external bodies is the processing of multisensory exteroceptive signals, such as vision and touch, which the central nervous system receives. I experience ownership over the fake hand because I feel touch on my hand and see touch on another hand at the same time. The prediction error generated in this situation is minimized by updating the hyperpriors (i.e., priors of a relatively high degree of abstraction; see Apps & Tsakiris, 2013; Clark, in press) that now produce a new generative model of which hand is more likely to be mine. The new generative model can now "explain away" the surprise by assuming ownership of the rubber hand, because different senses are weighted differently, with vision usually dominating touch. Interestingly, the way people describe such illusions is by referring first to sensations before talking about hands: "it seemed as if the touch I felt was caused by the touch I saw." This raises the intriguing question of whether what we actually own are the sensations evoked, rather than objects or body-parts. If it is the former that we own, then we should also try to understand the consciousness of the external body as a consciousness that consists in multisensory states, rather than objects. We would also then need to revisit the idea of "permanence of the body as an object" and think of the permanence of the body as a distinct *state* of sensory processing.

Looking for interactions between the internal and the external body

A recent study by Moseley et al. (2008) provides direct evidence that the experience of ownership of the external body, such as a rubber hand, is accompanied by significant changes in the homeostatic regulation of the real hand, which reflect at least some aspects that are typically involved in interoceptive processing. For example, skin temperature of the real hand decreased when participants experienced the rubber hand illusion (RHI). Additionally, the magnitude of the decrease in skin temperature on the participant's own hand was positively correlated with the vividness of the illusion. Importantly, this effect occurred only as a result of the experience of ownership. Thus, the experience of ownership, as a result of multisensory integration, over a new body-part has direct consequences for real body-parts, and for homeostasis (i.e., the physiological regulation of the internal body). Even more surprisingly, the RHI increases histamine reactivity, which is a key final pathway of the innate immune response, as measured by a limb-specific increase in the flare and wheal response to a skin-prick histamine test (Barnsley et al., 2011). These results suggest that the actual arm is being "rejected" and that the innate immune response is upregulated during the RHI.

To further investigate the interaction between the awareness of the body from *within* and from the *outside*, together with colleagues I combined an interoceptive sensitivity task with the same multisensory-based bodily illusion (the RHI) to test whether interoceptive awareness can predict the malleability of body-representations (Tsakiris, Tajadura-Jiménez, & Costantini, 2011). First, we measured interoceptive awareness with an established heartbeat-monitoring task (Schandry, 1981). We then quantified the extent to which participants experienced ownership over a fake hand using behavioral, autonomic and psychometric measures. Behavioral and autonomic measures of body-ownership malleability following exteroceptive stimulation were significantly predicted by interoceptive awareness, with low interoceptive sensitivity resulting in a stronger sense of body-ownership over a fake hand. These findings extend the interaction between exteroception and interoception observed by Moseley et al. (2008) in another direction. It seems that a stable perception of the body from the *outside*—what is known as "body image"—is partly based on our ability to accurately perceive our body from *within*, such as our heartbeat. The aforementioned results highlight the existing bidirectional interactions between the external and the internal body.

The id and the ego co-constitute self-specificity

Independently of whether we study the external body or the internal body, people experience "the feeling of the same old body always there," as James (1890, p. 235) aptly put it. However, the interpretation of this much-quoted phrase is ambiguous. In particular, is the "sameness" of the body to be taken literally here? Probably not. James refers to the continuing intimate experience of one's body to inform his analysis of personal identity. What is therefore meant here by "sameness" is the continuing presence of this intimate experience of one's body, rather than the continuing presence of a physically same or unchanging body. In other words, what is "always there" is not the same body, but the same quality of the experience of this body as *mine*. This intimate experience, reflecting the absence of predictions errors, would be based on efficient exteroceptive *and* interoceptive processing. Solms is right in proposing that the primary consciousness of the body is one that consists in states rather than objects, but my argument is that this should be extended to the external body and does not need to be restricted to the internal body alone. In the same way that interoceptive afferents reach the brainstem and produce changes in certain homeostatic set-points, multisensory integration, which occurs in the cortex, produces changes in sensory states. Both types of changes can be processed according to the same principles of free energy to produce the best-fitting model for explaining away surprises that are generated exteroceptively or interoceptively (Apps & Tsakiris, 2013). Given that surprises can occur as a result of exteroceptive or interoceptive information alone—or, even more interestingly as a result of an interaction, or even conflict, between the two types of information—the existence of a generative model that would processes self-specificity of the body at both levels would seem necessary. This generative model of embodiment, in the dual sense of the word (i.e., external and internal), would ensure that the brain can always use the best available exteroceptive or interoceptive data to resolve conflicts but, above all, to provide a continuing and a more or less coherent sense of being *me*, the *subject* of perception.

REFERENCES

Apps, M. A. J., & Tsakiris, M. (2013). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews.* [Epub ahead of print] doi:10.1016/j.neubiorev.2013.01.029

Barnsley, N., McAuley, J. H., Mohan, R., Dey, A., Thomas, P., & Moseley, G. L. (2011). The rubber hand illusion increases histamine reactivity in the real arm. *Current Biology*, 21 (23): 945–946.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13: 7–13.

Clark, A. (in press). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behavioural Brain Sciences*.

Freud, S. (1923). *The Ego and the Id*. *Standard Edition*, 19.

James, W. (1890). *The Principles of Psychology*. Cambridge, MA: Harvard University Press, 1981.

Merleau-Ponty, M. (1962). *The Phenomenology of Perception*, trans. C. Smith. London: Routledge.

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., et al. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences USA,* 105 (35): 13169–13173.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS ONE*, 3 (12), e3832.
Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, 18: 483–488.
Tajadura-Jiménez, A., Grehl, S., & Tsakiris, M. (2012). The other in me: Interpersonal multisensory stimulation changes the mental representation of the self. *PLoS ONE*. doi:10.1371/journal.pone.0040682
Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, 48 (3): 703–712.
Tsakiris, M., Tajadura-Jimenez, A., & Costantini, M. (2011). Just a heartbeat away from one's body: Interoceptive sensitivity predicts malleability of body representations. *Proceedings of the Royal Society of London, B. Biological Sciences,* 278 (1717): 2470–2476.

Facing inconvenient truths

Commentary by Oliver H. Turnbull (Bangor, UK)

This commentary offers a personal account of hearing Mark Solms's talk at the Berlin Congress in 2011 on the 'Conscious Id' concept. Organized around seven key thoughts, the commentary discusses the author's surprise at this radical idea; an opinion of its scientific basis; the likely impact of the idea on Freudian metapsychology; and questions around why this line of thinking had not been obvious years before and, especially, what might have been unique about that meeting in Berlin. In this regard, the effect of the simultaneous presence of Bud Craig, Jaak Panksepp, and Antonio Damasio is discussed, and especially the impact of Craig's radical position as a foil. Finally, the commentary considers Solms's bravery—for one so immersed in Freud—in admitting the "inconvenient" thought that Freud may, on this point, have been wrong.

Keywords: consciousness; Freud; history of science; id; metapsychology; neuropsychoanalysis

In the 15 years or so in which I have been associated with formally constituted neuropsychoanalysis, I have only twice experienced moments of genuine amazement at our annual conferences. The first was at the inaugural Congress of the International Neuropsychoanalysis Society in London in 2000. I had known (through my work with Mark Solms) of the concept of neuropsychoanalysis for much of my academic intellectual life. Indeed, the discipline has, for me, a prehistory of well over a decade. However, the London Congress was the first time that I realized, from the assembled group of hundreds, and the electric atmosphere, that there was a genuine community—spread across the globe—who appeared to be as passionate about neuropsychoanalysis as I was. It was a moment perhaps a little like a young music-lover first experiencing a nightclub after a childhood spent in the dull countryside.

I had to wait more than a decade for my second such experience, which arrived at our Berlin Congress in 2011. The moment came when Mark Solms stood up to offer the closing remarks after a session involving Bud Craig, Jaak Panksepp, and Antonio Damasio. As I recall, what Solms said (to paraphrase) was, "While I've been listening to these speakers talk about the role of various brain systems and emotion, it's made me think. And the conclusion that I've reached is that the emotional center of the mental apparatus, the part of the mind that Freud calls the Id, is quite clearly conscious! Indeed, it has now started me wondering about whether it isn't actually the *most* conscious part of the mind? Now, I'll have to think this through, because I've only just had this thought, and I realize that it runs counter to some core psychoanalytic ideas. But it feels more or less right."

The following day, in his concluding remarks to the Congress, Solms was able to outline his thinking on the topic more systematically. He began, as he always does, by peering at a few scribbled notes on a scrap of paper and saying, "I've only got a few simple things to say. So my speech shouldn't last more than five or ten

Oliver Turnbull: School of Psychology, Bangor University, Bangor, Gwynedd, UK.

minutes." He then laid out the argument that he had raised the previous day, together with an impressive sketch of the existing neuroscientific and psychoanalytic findings that would support this viewpoint. The Target Article in this issue is the written manifestation of that argument.

Why surprise?

And now to the question of my feelings of wonder. Through both of these episodes in Berlin, I sat astounded, my head reeling. A number of ideas, many of them novel as well as potentially incompatible, were regularly tossed in the air and reassembled. In retrospect, my principal thoughts were: (1) This idea was radically new and potentially revolutionary. (2) This idea was based on several really well-established scientific findings and seemed, in principle, correct. (3) This idea ran counter to a central tenet of Freudian metapsychology. (4) The delivery of this idea—by the current translator of Freud's *Standard Edition*—was psychoanalytic and journalistic dynamite. (5) Why had I not seen this obvious line of thinking before, given that I already knew most of the foundational material? (6) Why had Mark Solms not seen this obvious line of thinking before, because he (by definition) already knew all of the foundational material. (7) What was it about that morning in Berlin that had precipitated the idea in Solms's mind?

I left the auditorium and mentioned my moment of wonder to friends and colleagues. I recall using phrases such as, "The most important speech in psychoanalysis for 50 years"—referring, I think, to Points (3) and (4) above.

I will not seek, in this Commentary, to add to the issues raised in the Target Article (i.e., Points 1–3). The Target Article itself makes the issues of scientific support and theoretical importance very clear. However, I feel that I might be able to add some value to Points (6) and (7): the question of what it was precisely about that morning in Berlin that precipitated the idea, and why it was that Solms had not seen this before

Why not before?

In part, this is an old history-of-science question, most famously addressed by Thomas Kuhn in his *Structure of Scientific Revolutions* (1962). According to Kuhn's argument, most scientific work happens "in paradigm," filling the gaps in existing knowledge. Some findings seem incorrect or uncomfortable, but they are inconvenient truths, to be swept to the side while scientists concentrate on the more important business of confirming existing theory.

In Solms's case, the idea of a "conscious id" would have been a very inconvenient truth indeed. Much of his career has been invested in Freud scholarship: in doing psychoanalytic work, in translating Freud's writings into English; in gathering together Freud's preanalytic writings for (in many cases) their first-ever publication; and in working to synthesize psychoanalytic thinking with modern neuroscience (Kaplan-Solms & Solms, 2000; Solms, 1997; for review see Turnbull & Solms, 2007a). Importantly, much of Solms's career has also involved supporting Freud against his detractors, who come from both outside and within psychoanalysis (e.g., Blass & Carmeli, 2007; Hobson, 2007 [for a response, see Turnbull & Solms, 2007b]).

On the question of the veracity of psychoanalysis, Solms has always held the position that Freud may, in principle, be wrong on a number of points. However, his most common encounter with Freud-related questions has been to defend the founder of psychoanalysis against those who (often in misunderstanding) attack him. Thus, decades of protecting Freud may have meant that, even when Solms was peripherally aware of findings that were incorrect or challenging for psychoanalytic theory, these facts remained "uncomfortable," but not of sufficient danger to require a paradigm shift. They would, in Kuhn's view, be categorized as merely "inconvenient."

Why that day?

So much for the question of why this was not an obvious line of thinking *before* 2011. What, then, of Point (7)? What was it about that morning in Berlin? One line of argument might be the simultaneous presence of Craig, Panksepp, and Damasio. I cannot recall seeing at a neuropsychoanalytic event such an assembly of neuroscientists so interested and knowledgeable on the topic of the neural basis of emotion and consciousness. In addition, their positions were sharply polarized. Panksepp and Damasio are, in essence, staunch supporters of a subcortical, especially brainstem, basis for emotion and consciousness (e.g., Damasio, 2010; Panksepp, 1998). Importantly, they give emotion a place of primacy in both phylogenetic and ontogenetic terms, and each of them has laid out an impressive series of findings suggesting that the oldest, most primitive, parts of the mind form the central seat of emotion and the generation of conscious awareness.

In contrast, Craig offered an insula-focused claim about cortical emotion and consciousness (for a review, see Craig, 2009). His arguments were provocative, and they formed an extremely useful (and I suspect vital) foil to Solms's unfolding ideas.

Is it, then, that Solms had never heard others talk about the neuroscience of emotion and consciousness? Again, this cannot be the case, for he has heard Damasio talk on this topic before and has pored over various Damasio publications. In addition, he has heard Panksepp talk on the topic literally dozens of times.

Bud Craig as a foil

I would especially like to credit Craig's presentation, on the cortical basis of emotion and consciousness, with making the key difference. This is not—I suspect—because Craig brought directly useful and novel evidence to Solms but, rather, because his position was so provocative, offering a well-argued and evidence-based case for a radical position at odds with Panksepp and Damasio—and, indeed, Solms.

The position of a cortical role for conscious emotion has been presented before, often in relation to prefrontal cortex. However, this corticocentric argument is typically offered in an "enhancement" role, as organizer and facilitator of affective states (Davidson & Irwin, 1999; LeDoux, 1999), not as the core driver of conscious awareness. In contrast, Craig's "insula" variant of the cortical argument is a potentially more radical claim, given that it is based on anatomical data about that brain area as a hub for both internal and external perceptual sensation. The claim is for the insula as a (or *the*?) core source of emotional subjectivity. His work is also well referenced and is based on substantial research and a large literature (primarily in imaging and neuroanatomy).

However, as Solms knew very well, Craig's argument runs counter to a number of findings suggesting the importance of the brainstem (and related areas) for emotion and consciousness. For example, there are notable findings in animal (Panksepp, 1998), functional imaging (H. Damasio et al., 2000), and human-lesion (Parvizi & Damasio, 2003) research. Craig's claim also runs counter to findings showing the apparent unimportance of the cortex for generating emotion, based again on human-lesion research (Merker, 2007; Turnbull, Owen, & Evans, 2005) and on direct brain stimulation (Penfield & Rasmussen, 1950). These are clearly positions of strong contrast to that of Craig. Indeed, Craig's argument was so provocative that after the Berlin event of 2011, Damasio was himself inspired to immediately publish some new data (Damasio, Damasio, & Tranel, 2013), demonstrating that bilateral insular lesions produce no substantive change in conscious self-awareness—data that appear to further undercut Craig's argument.

Solms knew about these findings, though he had never before heard Craig speak. Now, in responding to Craig, he was at last forced to marshal his thoughts against a radical claim of the cortex as the generator of consciousness.

Freud's corticocentric approach

Now for the key moment. As mentioned above, Solms is used to translating arguments from neuroscience to psychoanalysis, and especially to Freud's metapsychology. A central point is that (as the Target Article makes clear) a corticocentric view of the type espoused by Craig is *exactly* the position held by Freud, who "locates the 'seat' of consciousness in the cerebral cortex" (Freud, 1923, p. 24). This placed Solms in a dilemma. On the one hand, he believes that the corticocentric position held by Craig cannot be reconciled with the available neuroscientific evidence. On the other hand, Sigmund Freud, the clinician–intellectual that Solms is so used to defending, is a supporter of that very position.

In such matters, it is always helpful to have some sort of defense, to protect us from the emotional consequences of holding such a counterintuitive position. In this case, the marginal defense is that Freud was (as are we all) a child of his time. His corticocentric position was simply the contemporary wisdom of the 1920s, with Freud himself noting that he had "merely adopted the views on localization held by cerebral anatomy, which locates the 'seat' of consciousness in the cerebral cortex" (Freud, 1923, p. 24).

It is often a bold move to choose the road less traveled. And it is clearly an act of bravery for one so immersed in Freud to admit that—on this point at least—Freud may have been wrong. Of course, it helps that the new position is backed by previously unappreciated evidence; however, one still has to *think* the inconvenient thought in the first place. Clearly, the evidence backing that thought had been available for years, and yet the idea had—with all its emotional baggage—remained elusive. That morning in Berlin cannot have hosted an easy decision. However, I believe that it was unequivocally the correct decision, for both psychoanalysis and neuropsychoanalysis, and I very much look forward to seeing whether this is an idea whose time has come.

REFERENCES

Blass, R., & Carmeli, Z. (2007). The case against neuropsychoanalysis: On fallacies underlying psychoanalysis' latest scientific trend and its negative impact on psychoanalytic discourse. *International Journal of Psychoanalysis,* 88: 19–40.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, 10: 59–70.

Damasio, A. (2010). *Self Comes to Mind.* New York: Pantheon.

Damasio, A., Damasio, H., & Tranel, D. (2013). Persistence of feeling and sentience after bilateral damage of the insula. *Cerebral Cortex*, 23 (4): 833–846.

Damasio, H., Grabowski, T, Bechara, A., Damasio, H., Ponto, L. L. B., Parvisi, J., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience,* 3*:* 1049–1056.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Science*, 3: 11–21.

Freud, S. (1923). *The Ego and the Id*. *Standard Edition*, 19: 12–59.

Hobson, J. A. (2007). Wake up or dream on? Six questions for Turnbull and Solms. *Cortex*, 43: 1113–1115.

Kaplan-Solms, K., & Solms, M. (2000). *Clinical Studies in Neuro-Psychoanalysis: Introduction to a Depth Neuropsychology*. New York: Other Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

LeDoux, J. (1999). Psychoanalytic theory: Clues from the brain. *Neuropsychoanalysis*, 1: 44–49.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30: 63–134.

Panksepp, J. (1998). *Affective Neuroscience*. New York: Oxford University Press.

Parvizi, J., & Damasio, A. R. (2003). Neuroanatomical correlates of brainstem coma. *Brain, 126*: 1524–1536.

Penfield, W. G., & Rasmussen, T. B. (1950). *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York: Macmillan.

Solms, M. (1997). *The Neuropsychology of Dreams*. Mahwah, NJ: Lawrence Erlbaum.

Turnbull, O. H., & Solms, M. (2007b). Big issues, little issues . . . and non-issues. *Cortex*, 43: 1116–1121.

Turnbull, O. H., Owen, V., & Evans, C. E. Y. (2005). Negative emotions in anosognosia. *Cortex*, 41: 67–75.

Turnbull, O. H., & Solms, M. (2007a). Awareness, desire, and false beliefs: Freud in the light of modern neuropsychology. *Cortex*, 43: 1083–1090.

Is the Primary Process Conscious?

Commentary by Gerald Wiest (Vienna)

Mark Solms's stimulating and thought-provoking article "The Conscious Id" provides new perspectives on our current understanding of concepts such as the mental apparatus, consciousness, or affects. Solms's ideas are challenging both for neuroscientists and psychoanalysts in many respects. In particular, he casts doubt on the corticocentric view of consciousness—that is, the neuroscientific assumption that all consciousness is cortical and that affective states generated deeper in the brain first have to undergo cortical processing before they become conscious. Psychoanalysts, on the other hand, are being confronted with the idea that the id—in contrast to the Freudian view—is the fount of consciousness and the ego is unconscious itself. In this commentary, I want to focus on different aspects of this proposed paradigm shift.

Keywords: condensation; consciousness; evolution; mentation; primary process

In "The Conscious Id," Mark Solms makes us aware of a widely unrecognized fact about the Freudian model of the mental apparatus—namely, that Freud conceptualized consciousness as an exclusively cortical process. The origins of this view may be traced back to the hierarchical-evolutionary theories of Herbert Spencer, who influenced Freud in his metapsychological concepts. Spencer (1855) posited that the mind, like all structures in the universe, develops from a simple, undifferentiated, homogeneity to a complex, differentiated, heterogeneity, while being accompanied by a process of greater integration of the differentiated parts. In this view, human consciousness and self-awareness are implicitly localized in the evolutionarily youngest part of the brain—the neocortex. Spencer also envisioned that during phylogenetic development, older sites of consciousness are subsequently being transformed into reflex-centers, which are then no longer accessible to higher developed sites ("Beyond the limits of the coherent aggregate of activities . . . constituting consciousness, there exist other activities of the same intrinsic nature, which being cut off are rendered foreign to it"; Spencer, 1855).

Spencer's theory of an evolutionary organization of the brain had a significant impact on John Hughlings Jackson's understanding of brain function (Hughlings Jackson, 1884). According to Hughlings Jackson (who in turn influenced Freud in his conceptions of the mental apparatus), the brain evolves through an increasingly complex coordination of different brain modules or units, where every higher level re-represents its subordinate levels. Thus, he conceived the cortex—representing the highest center for integration and coordination—as the seat of consciousness and self-awareness. The influence of the hierarchical brain theories of Spencer and Jackson are evidently demonstrated in Freud's graphical models of the mental apparatus, showing the system of perception-consciousness (Pcpt.-Cs.) always at the top of the figure (corresponding to the cortex).

The hypothesis that the id is intrinsically conscious is a challenge for traditional psychoanalytic models of the mind. The Freudian id is usually understood as a mental structure that corresponds to what previously had been encompassed by the concept of the "unconscious." It is regarded as the seat of the instinctual drives, and it functions according to the pleasure principle and the primary process. The idea that the id is a conscious agency not only represents a paradigm shift in psychoanalysis but may also provide answers to previously unresolved questions. The question that instantly comes up after reading Mark Solms's article is: how can the id as a hallucinatory, unrealistic, not time-bound, and irrational agency by nature represent a conscious entity? However, the id also exhibits some characteristic features that are not fully compatible with a completely unconsciously operating structure.

One of the main operational hallmarks of primary-process mentation is, besides displacement, the mechanism of condensation (i.e., two or more different elements are combined to form a new one). Freud himself was fascinated by the capacity of the primary process by describing condensation as an elaborate technique that makes use of occasional similarities between two objects in the most intelligent way. Condensation is, for example, a key element in dream formation. Freud refers to the phenomenon in dreams by which many ideas may be condensed into a single image. A single character in a dream may, in the form

Gerald Wiest: Department of Neurology, Medical University Vienna, Vienna, Austria.

of a "composite figure," be identified by the dreamer as representing different people at the same time.

It is interesting to note that Freud held fast to the energetic model of the primary process, even after having introduced the structural theory of the mental apparatus. This energetic view was later criticized by some authors, as the primary process has not been conceptualized from a structural perspective. Holt (1967) emphasized that the existence of "function" always also necessitates the existence of "structures." In dreams, condensation is always characterized by an amalgamation of images or mental contents that share common features, which implies that they are not being selected randomly. It is exactly this "unrandomness" that calls for an organizing or structuring component within the primary process. A look at recent visual computer techniques might help to understand the organizing mechanisms being effective in condensation. The technique of a "photographic mosaic" (http://en.wikipedia.org/wiki/Photographic_mosaic) is a specific computer application that uses content- or color-related similarities of photos to create a new image in terms of a mosaic. In a way, this mosaic resembles dream images (of characters, places, etc.) or contents, which are also condensed and fused into one object. The computer mosaic is created by means of a specific software—that is, an algorithm (= structure)—that adjusts hundreds of images according to their similarities. Close examination of the resulting mosaic reveals that the image is obviously made up of many hundreds or thousands of randomly selected smaller images (similar to the seemingly chaotically mixed images of manifest dream content), but when viewed from a distance, a new image emerges (similar to the latent dream content). The fact that the primary-process-associated mechanism of condensation can be simulated by a computer algorithm is, in my opinion, a clear indicator that the primary process as the functional principle of the id is not a chaotic, uncontrolled, or randomly operating system, but may be conceptualized as a specific information-processing system that acts according to synthetic requirements (Holt, 1967). If the primary process is indeed capable of applying, as Freud has put it, an elaborate technique that makes use of occasional similarities between two objects in the most intelligent way, then one has to assume that the id possesses organizational capacities, which implies that the id is—at least to some extent and consistent with Mark Solms's proposal—conscious.

In his article, Solms refers to a condition called hydranencephaly to confirm that the corticocentric view of consciousness is mistaken and that human beings can be conscious without having a cortex. This clinical example strongly supports the concept of a conscious id. However, Solms does not provide any suggestions as to how his hypothesis can be tested further empirically. After all, it should be in the interest of psychoanalysts to gain experimental proof of their theories and hypotheses. As in all sciences, assumptions are supposed to be tested using methods independent of those that presuppose the assumptions to be true (Brakel, 2004).

Primary-process mentation has already been tested experimentally in numerous studies, showing that complex and meaningful psychological operations take place on content that is unconscious (Brakel, Kleinsorge, Snodgrass, & Shevrin, 2000; Shevrin, 1992; Snodgrass, Shevrin, & Kopka, 1993; Wong, Bernat, Bunce, & Shevrin, 1997). These experimental data suggest that the primary process may be of even more importance than Freud suspected. In this regard, it has been hypothesized that primary-process mentation represents the basic mental organization in many nonhuman mammals (Brakel, 2004). However, if we accept the idea that the primary process is the major mode of thought organization operative in nonhuman primates or other higher mammals, then the idea of a "conscious id" would be conceivable as well. Thus, in accordance with Freud one is tempted to recognize the primary process as an archaic form of thinking or consciousness, in the sense of a phylogenetic atavism. If the primary process is conceived as a phylogenetically conserved archaic mental process, then what is its evolutionary significance? The ethological concepts of so-called fixed-action patterns and sign stimuli may provide a key to the understanding of the adaptive role of the primary process. Fixed-action patterns represent the simplest type of behavior, in which a specific stimulus (a "sign stimulus") nearly always results in an invariable behavioral response. Nikolaas Tinbergen (1951) discovered that fixed-action patterns can also—and even more easily—be elicited by the application of symbolically abstracted sign stimuli (a wooden stick with a red dot on the end evokes the same, or even enhanced, pecking response from nestlings as does the red spot on the beak of a herring gull). The operative mechanisms in sign stimuli may thus be similar to displacement, condensation, or symbolization—the basic principles of primary-process mentation. One of the main purposes of primary-process mentation could thus be the activation of the motivational system and the induction of a specific behavior. The capability of symbolic transformation could, from an evolutionary perspective, be understood as an adaptive mechanism to increase the redundancy of the sign stimulus (Wiest, 2010).

In his "triune brain theory," Paul MacLean proposed similar concepts of archaic mental functions, which are preserved across all mammals including humans. MacLean divides the triune brain functions into three types of mentation: protomentation, emotomentation, and ratiomentation. Protomentation in the reptilian brain refers to prototypical behavior or mental states such as drives and impulses. Emotomentation in the mammalian brain refers to emotion-related perceptions and behavior. Ratiomentation, as an exclusive feature of the neocortex, refers to mental activity related to thinking (MacLean, 1990). MacLean, though not proposing as Solms that the neocortex is itself unconscious, implicitly uses the term "mentation" for mental activities in archaic brain structures, such as the basal ganglia (ratiomentation) or the limbic system (emotomentation), which attributes at least some kind of consciousness to these deep-brain structures (as does Solms to the id).

Solms's article is not only a remarkable attempt to revise the traditional and sometimes rigid concepts of psychoanalysts and neuroscientists; his new concepts may also provide answers to hitherto mostly neglected questions in these disciplines. However, while Solms's challenging ideas have to await empirical confirmation, it has to be considered that a change of perspective is sometimes essential for the development of new directions in science, or, as Albert Szent-Györgyi (1957, p. 57) might have put it: discovery is seeing what everybody else has seen, and thinking what nobody else has thought.

REFERENCES

Brakel, L. A. (2004). The psychoanalytic assumption of the primary process: Extrapsychoanalytic evidence and findings. *Journal of the American Psychoanalytic Association,* 52: 1131–1161.

Brakel, L.A., Kleinsorge, S., Snodgrass, M., & Shevrin, H. (2000). The primary process and the unconscious: Experimental evidence supporting two psychoanalytic presuppositions. *International Journal of Psychoanalysis*, 81: 553–569.

Holt, R. R. (1967). The development of the primary process: A structural view. *Psychological Issues,* 5 (2–3): 344–383.

MacLean, P. D. (1990). *The Triune Brain in Evolution: Role in Paleocerebral Functions*. New York: Plenum Press.

Shevrin, H. (1992). Subliminal perception, memory, and consciousness: Cognitive and psychodynamic perspectives. In: *Perception without Awareness: Cognitive, Clinical, and Social Perspectives*, ed. R. Bornstein & T. Pittman. New York: Guilford Press, pp. 123–142.

Snodgrass, M., Shevrin, H., & Kopka, M. (1993). The mediation of intentional judgments by unconscious perceptions: The influences of task strategy, task preference, word meaning, and motivation. *Consciousness & Cognition,* 2: 169–193.

Spencer, H. (1855). *The Principles of Psychology*. London: Longman, Brown, Green and Longmans.

Szent-Györgyi, A. (1957). *Bioenergetics*. New York: Academic Press.

Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Clarendon Press.

Wiest, G. (2010). Die psychoanalytische Theorie des Primärprozesses. *Psyche,* 64: 212–235.

Wong, P. S., Bernat, E., Bunce, S., & Shevrin, H. (1997). Brain indices of nonconscious associative learning. *Consciousness & Cognition*, 6: 519–544.

Where Ego Was, There Id Shall Be? Some Implications of the "Conscious Id" for Psychodynamic Psychotherapy

Commentary by Maggie Zellner (New York)

Mark Solms's perspective on the "conscious id" posits that brainstem nuclei and cortex can be mapped, in broad brush strokes, to subject and object: brainstem nuclei mediate the core experience of being, where perception, emotion, and action are integrated at a primal level, while the cortex elaborates on representations of the contents of consciousness. In this commentary, I expand on some of the intrapsychic implications of this distinction, particularly the dynamics of disproportionately perceiving oneself as an object (seeing one's self through the eyes of the other), at the expense of experiencing being a subject, and seeing the world from one's own perspective. Solms also correlates the activity of cortical circuits that mediate learning, predicting, and automating procedures with ego functions that serve to constrain primary consciousness. I discuss some of the implications of this perspective for psychodynamic treatment, which aims to update certain maladaptive processes by bringing affective awareness to previously automated processes.

Keywords: brainstem nuclei; cortex; object; psychodynamic treatment; subject; third-person perspective

The best things in life are experienced consciously—the fun of play, the satisfaction of achievement and connection, the pleasure of sex, and so on. However, because of rules internalized during our early experience, which often operate largely out of awareness, some of us who are lucky enough to have decent living circumstances still have difficulty living as fully as possible. This is what brings many people to therapy—the conscious and unconscious processes that keep us from creating the conditions to have more pleasure, engagement, and fulfillment. Mark Solms's Target Article presents a framework for a revised neuropsychoanalytic model of the brain and mind, which raises numerous points deserving commentary, critique, elaboration, and testing. In this commentary, I focus on some of the clinically relevant aspects of his proposal.

The brainstem is critical for *being a subject* with consciousness, the cortex for *representing the objects* of consciousness

> The relationship between the two aspects of consciousness—the objects and the subject of perception—is also what binds the component of perception together; objects are perceived by an experiencing subject (cf. the "binding problem"). [Solms, Section 1]

In broad brushstrokes, Solms correlates subject and object with the subcortical and cortical levels of the brain, respectively. Obviously, the levels of brainstem and cortex are not easily separable in an intact adult brain, being connected in numerous interacting and reciprocal circuits; at the level of intrapsychic functioning and subjective experience, subject and object are similarly deeply interwoven. But I agree that the evidence does indicate that a rough division is justified: the brainstem nuclei such as the periaqueductal grey and parabrachial nucleus fundamentally integrate sensory, motor, and affective processes, constituting the "protoself" (Damasio, 2010), while the cortex appears to mediate the more highly elaborated levels of representation, integration, and regulation that constitute or constrain the specific contents of consciousness. In Solms's formulation, the brainstem is therefore critical for generating the mental space of the *subject*—with an experience of *being*—who can attend to particular objects of consciousness. (Here we are in the company of leading scientists who signed the recent "Cambridge Declaration of Consciousness," stating that "The absence of a neocortex does not appear to preclude an organism from experiencing affective states"; Low, 2012.)

Solms's model may help illuminate the neural processes of a central dynamic in low self-esteem, depression, narcissistic vulnerability, and social anxiety: the problem of seeing one's self through the eyes of the other, and therefore seeing one's self as an *object*, rather than living more as a *subject*. As social creatures, we devote significant resources to attending to others and to generating a theory of mind of their intentions toward us. This is adaptive and is an integral part of being human and caring about others. However, when this is out of balance, we care more about how the other perceives us than what our own perceptions are—for example, self-consciousness at a cocktail party interfering with making small talk, when concern about *appearing interesting* gets in the way of

Margaret R. Zellner: The Neuropsychoanalysis Foundation, New York, NY; & The Rockefeller University, New York, NY, U.S.A.

feeling interested. More pathological examples include procrastinating for fear that others will criticize or reject, or engaging in self-destructive behavior to comply with the abusive behavior of a significant other (or with the demands of an introject).

I hypothesize that devoting disproportionate resources to representing the other's image of us, rather than maintaining our own perspective, could produce dysfunction in at least two ways. First, our mental representations of a disapproving, rejecting, or hostile other (mediated mainly by cortical circuits) generate emotional responses such as fear, separation distress, or aggression (mediated by limbic and brainstem nuclei). These affects then require various mechanisms of affect regulation (instantiated mainly at the cortical level), dampening our overall emotional responsiveness. This could be one piece of the puzzle in depression, for instance (which Freud, 1917, famously described as the "shadow of the object falling upon the ego"), since depression is often marked by hyperactivation of the ventromedial prefrontal cortex (Mayberg et al., 1999), a region implicated in affective regulation (Urry et al., 2006).

Second, disproportionately taking the other's perspective on one's self might lead to undue inhibition of one's own basic motoric and motivational apparatus, perhaps through specific mechanisms involved with perspective taking. First-person and third-person perspectives engage overlapping as well as distinct brain areas (e.g., see Ruby & Decety, 2001), and navigating the world through a third-person perspective may be unwieldy. To my knowledge, this hypothesis has not been directly tested, but some converging evidence supports the idea. For example, perceiving objects from a first-person perspective induces a "disposition to act," correlated with differential activation of the intraparietal sulcus compared to the third-person perspective (Kockler et al., 2010); reading first-person sentences activates the caudate to a greater extent than third-person sentences (Otsuka, Osaka, Yaoi, & Osaka, 2011); and third-person memories are associated with deactivations of insula and somatomotor areas (Eich, Nelson, Leghari, & Handy, 2011). I have the clinical sense that living one's life as an object rather than a subject siphons off some of the basic "energy" for living, as if seeing one's self from the point of view of the other suppresses the SEEKING system, leaving one with a deficit in energy for moving out into the world and taking action.

I think many therapists would agree that psychodynamic treatment works on shifting this balance when it is out of whack—facilitating the ability to feel one's *own* feelings and desires, and recognizing our reactions to what we imagine *others* feel or perceive (and fundamentally realizing that they are *our* fantasies). The forces of the id—drives, motivations, affective responses—are motor/sensory programs that animate us and are correlated with a sense of aliveness in the here-and-now, arising from an experience of our own body (particularly the "internal" body, as Solms stresses), a taking of our own perspective. Therefore, when we bring the focus there, we have more access to our capacity to act. In contrast, being too involved with our representations of the other's representations of ourselves leaves us without much phenomenological ground to stand on, as it were.

Learning, predicting, and automatizing: the "unconscious ego" suggests an overlap between the dynamic unconscious and the cognitive unconscious

I agree with Solms that the neuroscientific evidence suggests that the ego is largely unconscious, mediated by cortical circuits whose operations are devoted in various ways to learning, automatizing, and predicting—making the conscious unconscious, as it were. Arlow (1969) likened the effect of unconscious fantasy on perception (correlated in this context with our early templates or automatized predictions, which constrain awareness) to the process of two films being projected on a movie screen at once. If our unconscious fantasies or expectations are significantly out of sync with current reality, affects are mobilized or inhibited inappropriately and mental resources are recruited to keep content out of awareness or to filter our perceptions. In this regard, what is problematic for many of us are the *maladaptive* automatized processes that interfere with being able to respond appropriately to the world and act on our inner needs. (I emphasize "maladaptive" since automatization, in and of itself, certainly is not a problem—it allows us to engage productively with the world.) Therefore, understanding the neural mechanisms of these processes is key to moving forward in treatment and diagnosis.

In his Target Article, Solms takes an important step: he notes (almost in passing) that the correlation of the processes of learning, regulation, and automatization with aspects of the dynamic unconscious indicate that our prior separation of the dynamic unconscious and the "cognitive unconscious" should be dismantled, because there is apparently a great deal of overlap to be explored. This is a nice step forward in the dialogue between neuroscience and psychoanalysis, given the wealth of operationalized measures and prior findings in the realm of the cognitive unconscious (see the Target Article by Berlin, 2011, in a previous issue of this journal), which may now be fruitfully exploited in further experimental work.

Where ego was, there id shall be?

> The therapeutic task of psychoanalysis, then, would still be to undo repressions (to allow the associative links to regain episodic status) in order to enable the reflexive subject to properly master the object relations they represent and generate executive programs more adequate to the task, so that they may then be legitimately automatized. [Solms, Section 10]

If maladaptive automatized processes are a central dynamic in psychopathology, then syncing them better with the here-and-now allows us to engage more appropriately with the world. Bringing into awareness previously excluded thoughts and feelings—one of the ways of "re-syncing" automatized processes—is one of the core features of psychodynamic treatment, and Solms's formulation is a platform for further exploring the neural mechanisms of the process and why it facilitates change.

I will mention just a few processes that may be relevant here that deserve further exploration. First, bringing attention to something that has been transferred to procedural memory can disrupt performance (Beilock, Bertenthal, McCoy, & Carr, 2004), perhaps by changing prefrontal inputs to striatal circuits. This may have relevance in psychoanalytic treatment to bringing attention to expectations and patterns of behavior that are taken for granted, including working in the transference when it is "hot." Second, bringing automatized processes into awareness may involve creating new patterns of integrated activity between various networks in the brain, including the attentional networks, the default mode network, the salience network, and the limbic networks. Finally, conscious, affective awareness of templates—when they are "online"—in the new emotional context of psychodynamic treatment may facilitate reconsolidation at the synaptic level (Schiller et al., 2010), allowing for the creation of new pathways and the sculpting of old ones.

Hopefully, as we develop "executive programs more adequate to the task," as Solms proposes, we come to expect a more benign environment, increase our ability to tolerate loss, feel less existentially dependent on our ties to significant others, and so on. We can then enjoy the use of our cortical processes—our capacity for planning, creativity, detailed perception, taking action—to a greater extent.

Conclusion

As children, when feelings or behaviors lead to negative reactions in our significant others, we learn to avoid thinking, feeling, or acting in certain ways, often through the activation of automatized mental processes. The "conscious id" perspective is a framework for exploring the brain-mind mechanisms by which bringing these maladaptive processes into awareness helps transform them. Resolving these processes—grieving losses, revising negative identifications and self-representations, softening overly negative or overly idealized expectations of others, and so on—creates the conditions for fuller expression of the primary motivational and affective processes that constitute the meaningfulness of life. In doing so, psychoanalytic treatment ideally helps us to be more fully present—to live more from the brainstem, as it were.

I do not advocate living purely "in the id," of course (if that were even possible with an intact central nervous system)—we would be limited mainly to responding to here-and-now stimuli, obeying the pressure to satisfy impulses as soon as they arise, unable to plan or regulate our behavior. Indeed, we depend on ego processes that constrain experience so that we can think, and so that we can regulate our behavior to maximize connection and cooperation with others. However, we need not be afraid of the id, as if our only pleasure arises from rampaging through the world satisfying our selfish needs—our understanding of affective neuroscience demonstrates that we have drives to relate to objects, to play with others, and to care about them (Panksepp, Nelson, & Bekkedal, 1997). Since we take primary pleasure in playing with and caring for others, our id therefore fuels a motivation to use our cortical processes to regulate our behavior in a way that brings pleasure and avoids harm to others.

But we want to be regulated without being constricted. We don't want defensive processes to interfere with pleasure, or inhibit our striving for connection and achievement for fear of loss or failure. We don't want to relate to internal fantasies at the expense of real relationships, or to live in the past or project into the future at the expense of the here-and-now. And we don't want to let empathic attunement for others obscure our own emotional realities and significantly impair our own aliveness. Analytic treatment allows for progress in all of these domains, in a powerful way. Solms's perspective on the "conscious id" contributes to our exploration of the neural and intrapsychic mechanisms of these processes.

REFERENCES

Arlow, J. A. (1969). Unconscious fantasy and disturbances of conscious experience. *Psychoanalytic Quarterly*, 77 (1): 21–46.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review*, 11 (2): 373–379.

Berlin, H. (2011). The neural basis of the dynamic unconscious. *Neuropsychoanalysis*, 13 (1): 5–31.

Damasio, A. R. (2010). *Self Comes to Mind: Constructing the Conscious Brain.* New York: Pantheon.

Eich, E., Nelson, A. L., Leghari, M. A., & Handy, T. C. (2009). Neural systems mediating field and observer memories. *Neuropsychologia*, 47 (11): 2239–2251.

Freud, S. (1917). Mourning and melancholia. *Standard Edition,* 14.

Kockler, H., Scheef, L., Tepest, R., David, N., Bewernick, B. H., et al. (2010). Visuospatial perspective taking in a dynamic environment: Perceiving moving objects from a first-person-perspective induces a disposition to act. *Consciousness and Cognition*, 19 (3): 690–701.

Low, P. (2012). *The Cambridge Declaration on Consciousness*. [Statement released at the Francis Crick Memorial Conference on Consciousness in Human and Non-Human Animals, at Churchill College, University of Cambridge.] Available at: http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., et al. (1999). Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. *American Journal of Psychiatry*, 156 (5): 675–682.

Otsuka, Y., Osaka, N., Yaoi, K., & Osaka, M. (2011). First-person perspective effects on theory of mind without self-reference. *PLoS ONE*, 6 (4), e19320.

Panksepp, J., Nelson, E., & Bekkedal, M. (1997). Brain systems for the mediation of social separation-distress and social-reward: Evolutionary antecedents and neuropeptide intermediaries. *Annals of the New York Academy of Sciences*, 807: 78–100.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience,* 4 (5): 546–550.

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463: 49–53.

Urry, H. L., van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., et al. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience*, 26 (16): 4415–4425.

Response to Commentaries

Mark Solms

THE ID IS NOT THE SAME AS THE UNCONSCIOUS . . . AND OTHER THINGS

I am of course pleased by the substantial degree of agreement with my main conclusions. Due to limitations of space, I will focus on the points of disagreement.

Ariane Bazan says I have conflated affect and drive (i.e., *Trieb*, which Strachey called "instinct"). She also thinks that I confuse drive and instinct, in the modern sense of the word. First, let me say that I did not intend to do so, because I do not consider these things to be synonymous at all. I define drive as "a measure of the demand made upon the mind for work in consequence of its connection with the body" (Freud, 1915a, p. 122), where the "measure" is the degree of deviation from a homeostatic set-point (with implications for survival and reproductive success).[1] I do not believe that this deviation itself is something mental, but the "demand" it generates is felt in the pleasure–unpleasure series. This (felt demand) is affect, which in my view is the origin of mind. The "work" that flows from affect is cognition, the functional purpose of which is to reduce affect—that is, to reduce prediction error (free energy). The purpose of cognition is to bring the

[1] A general point that applies to all the commentaries: When we use the word "body" we generally think of our exteroceptive (mainly visual) representation of it. This is not strictly correct. Our visual image of the body—both its external surface and the viscera—is no more real than our affective sense of it. The ubiquitous tendency to privilege the visual (and other exteroceptive senses) produces the mind–body problem. It leads us to the erroneous impression that the visually represented body *causes* affective feelings, when in fact the two apparent things—the body-as-object and the body-as-subject—are simply two different ways of perceiving the same thing. (See Solms, 1997; Solms & Turnbull, 2002.)

world into line with our predictions and our predictions into line with the world. This work centrally involves learning. Learning would be unnecessary if the solutions to life's problems were preordained. That is why the objects and even the aims of drives are typically not preordained—they have to be learned. The exception to this rule is instinct (in the modern sense). Instincts obviate the need for cognition (apart from conditioning of "adequate" objects). Each instinct responds in its own stereotyped (unconditioned) fashion to its adequate objects—some of which may be innate (cf. Bion's "pre-conceptions")—on the basis of inherited predictions (what Friston calls "full priors"). Freud called such predictions *primal phantasies*, which are "very early impression[s] of a very general nature, placed in the prehistory not of the individual but of the species" (1916–17, p. 395). I hope it is clear why I do not consider drives to be instincts. The feelings that are evoked by the adequate stimuli of instincts classify (or "explain") the unpredicted events that evoke them in accordance with "some particular significant experience . . . placed in the prehistory not of the individual but of the species" (p. 395) and, accordingly, trigger the relevant inherited (unconditioned) response. It is important to add that instinctual feelings—feelings associated with the "basic emotions"—are intrinsic to the brain; they are not current perceptions of "oscillations in the tension of its instinctual needs [*Triebe*]" (Freud, 1940, p. 198). As far as instincts are concerned (as opposed to drives), Bazan is therefore wrong to suggest that "*the pleasure criterion is given by the (internal) body*—not by the brain." This is why it is so easy to trick the brain, as occurs with all psychotropic drugs. Heroin, for example, does not cause pleasure "only if the motor pattern chosen is successful" in removing the actual internal bodily need (Bazan). Affect cannot be reduced to perception of the actual state of the body. I do not subscribe to the James–Lange theory. Before leaving Bazan's commentary, I must also disagree with the way she equates inhibition and repression. Inhibition has a broader functional ambit than repression. Thus it frequently happens that an instinctual emotion is inhibited without being repressed; the subject may remain conscious of the primary inhibited response and even use it to guide the secondary cognitive one (cf. Freud's 1926 concept of signal anxiety).

In stark contrast to the other commentators, **Heather Berlin** says that my claim that affective consciousness can occur in the absence of cortex and that consciousness cannot occur in the absence of brainstem are "unsupported assumptions." Considering that I summarize several lines of evidence for these conclusions, Berlin's claim would be incomprehensible were it not for the fact that she defines consciousness differently from me. What I call affective consciousness she (like most cognitive neuroscientists) calls *wakefulness*. She agrees that wakefulness is generated from upper-brainstem nuclei, which she characterizes as "the power supply to the brain"; however, in her view these nuclei are mere "enabling factors" of consciousness. What Berlin (again, like most cognitive neuroscientists) calls consciousness is *awareness*, which she characterizes as the "*content* of consciousness." She argues that "specific factors" are required to generate awareness of "any one conscious percept"—namely, short-term correlations between populations of corticothalamic neurons. Since I agree that this (exteroceptive) type of awareness requires cortex, the difference between Berlin and me appears to boil down to the question as to whether affect should be defined as "content" or not (or whether *it feels like something* to be awake or not). But Berlin goes further: she argues that affect (feeling like something) is not generated from upper-brainstem nuclei. Since this is an empirical rather than a definitional question, Berlin has to grapple with the evidence I cited in support of the other view. Unfortunately, however, she grapples with it primarily on philosophical grounds. She claims that the evidence for my view that hydranencephalic children (and decorticate animals) display affective consciousness "begs the question entirely" because this evidence entails the "moralistic fallacy" (a belief that *what ought to be* corresponds with *what is*):

> . . . showing that strongly conserved emotional facial displays and conditioned responses from a brainstem/spinal-cord system can be developed over time says nothing about whether conscious emotional states attach to these observable phenomena. . . . We cannot *assume* that having a sleep–wake cycle and expressions of emotion (laughter, rage, etc.) necessitates consciousness. For example, we can produce similar pseudo-emotional reactions in nonconscious machines. . . . While it is true that they may in fact be consciousness, we cannot assume that they are. . . . The crux of Solms's theory relies on a projection of the existence of consciousness based on what *look like* meaningful emotional behaviors. [Berlin; emphasis added]

Readers will notice that Berlin here equally *assumes* the opposite: that the laughter, rage, etc., of hydranencephalic children and decorticate animals are "pseudo-emotional." This might be called the *behaviorist fallacy*, which places higher demands on evidence for consciousness than any other type of scientific evidence. Why should we assume that contextually appropriate emotional displays, which are readily evoked by stimulation of a particular brain region and obliterated by lesions of that same brain region and which correspond to affective feelings in ourselves, do *not* correspond to affective feelings in these children and animals? Surely that assumption would be more "arbitrary" (Berlin) than mine. The only evidence for it is that these children and animals cannot "declare" their feelings *in words*. In this connection, Berlin would do well to remember that declarations of feeling do not prove their existence; even nonconscious machines can be programmed to "declare" feelings. I do not even know for sure whether Heather Berlin herself experiences feelings: one can only directly confirm the existence of one's own feelings. This is the philosophical problem of "other minds"—the burden of all psychology, not of my article. As Berlin says, we simply "have no way of knowing." The only way out of this conundrum is the ordinary scientific method. If behavioral predictions arising from the hypothesis that affective feelings are generated in the upper brainstem are confirmed experimentally (as they always are), then this hypothesis must be accepted, *unless and until positive evidence for the opposite thesis is produced*. To proceed otherwise is to render impossible a science of consciousness. But Berlin goes further: even if the presence of consciousness in hydranencephalic children *is* positively proven, she still will not accept it. Due to neuroplasticity, she argues, the brainstem in these children might have sprouted "cortical" functions. In the remainder of her commentary, Berlin cites various bits of evidence that suggest that upper-brainstem lesions do not *necessarily* or *always* or *completely* obliterate consciousness, unless the lesions are "extensive, bilateral, and extend rostrally." Well, precisely. It is striking how Berlin's standards of evidence shift, and how she gives all the benefit of the doubt to the cortex, and thus arrives at the conclusion that cortex without brainstem *just might* be capable of consciousness. Readers left confused by all this should consult Parvizi and Damasio's (2003) authoritative study. Coma (loss of wakefulness *and* awareness) is reliably associated with focal bilateral tegmental lesions of the upper pons and midbrain—and that's that.

Robin Carhart-Harris does not dispute the fact that a "primitive sort" of consciousness is generated in the upper brainstem; but he questions its relevance to the Freudian model of the mind, "which depends on the notion of conflict between the ego and the id" (Carhart-Harris). Second, he questions my supposed localization of the ego in the sensorimotor convexity, arguing instead for the midline default mode network (DMN). Third, although he accepts that "there seems no reason to doubt that [hydranencephalic children] possess a rudimentary form of consciousness, with a varied emotional repertoire," he points out that consciousness is also driven from deep thalamic regions that remain intact in these children. Fourth, he reminds us that brainstem nuclei are part of an interconnected system. He thus argues that the id should be equated with the extended limbic system, not with the upper brainstem. He concludes that a focus on the limbic system and its relations with the DMN "may be [a] too general and unspecific" way of characterizing the relationship between id and ego than my focus on brainstem and cortex. These are fair comments. My article emphasizes the extremities of the id and ego: the autonomic core and the sensorimotor periphery, respectively. However, more interesting interactions certainly occur in the overlapping zones, where these two poles of the mind must be reconciled. (Incidentally, I include the so-called nonspecific thalamic nuclei in the upper brainstem.)

In a similar vein, **Katerina Fotopoulou** fears that my "denigration" of the id to affective consciousness (see my response to Bazan) and of the ego to mere representation and automatization "risks de-emphasizing the central place of the Freudian antagonism between an inflexible body and an unpredictable world" (Fotopoulou). As with Carhart-Harris, I find myself in agreement with most of what Fotopoulou says, and I am not sure why she thinks I might not. I will, however, make two points. When I theorized the cortical ego as "the driver of automaticity" (Fotopoulou), I did not mean to imply that the ego lacks flexibility in relation to the changing world; what I meant was that it *aspires to* reduce its need for flexibility (it aims to *master* this unpredictable world). That is the ego's *raison d'*être. In this regard, the ego ultimately serves the interests of the id, although it is better equipped than the id to cope with the outside world. Second, I agree with what Fotopoulou says about the dichotomy between uncertainty and precision. Precision is probably the defining feature of cortical consciousness. But I would equate uncertainty with affect, precision (reliability of predictions) with cognition. I am therefore not sure what she means when she says that affect monitors uncertainty rather than hedonic quality. To my way of thinking, hedonic quality *is* our measure of uncertainty (of free energy). What is reliably predictable is neither pleasurable nor unpleasurable; it is boring. Hence, as Fotopoulou points out, the "ultimate guiding principle" of the mind is not pleasure, it is Nirvana, which lies "beyond the pleasure principle" (Freud, 1920). What is new in this respect is the fact that the *ego* seeks Nirvana. The id is constantly "surprised" by reality, with all its unexpected (unwanted) constraints—hence its endless feelings and fuss. The ego, by contrast, becomes progressively older and wiser. To quote Fotopoulou: "It thus falls upon the ego—or cognition—to tailor this inflexible, inherited minimization imperative to the demands of the unpredictable world during one's life-time. Under perceptual and active inference. the ego thus builds empirical priors on the foundations of innate priors." Precisely—I agree!

Karl Friston's dense and deep commentary repays multiple re-readings. He concurs that ego cognition seeks to reduce the need for flexibility, to minimize prediction errors and maximize precision: "The binding of free energy (prediction errors) corresponds to a top-down suppression, which necessarily entails an explanation or resolution of violated predictions." Somewhat like Carhart-Harris, he points out that the "top" of this hierarchy is not the sensorimotor cortex—and nor, of course, the autonomic core—it is the overlapping zones *between* internal and external. "Put another way, high-level intransigent representations (mental solids) have an amodal aspect and provide bilateral top-down interoceptive *and* exteroceptive predictions" (Friston). I am, however, unsure I agree with Friston when he associates consciousness with prediction-making "probabilistic representations," notwithstanding the fact that he accepts that *intrinsic* consciousness (intrinsic prediction or "full priors") is generated from the brainstem and limbic system. To my way of thinking, consciousness is associated with violated predictions ("surprise," free energy), not with predictions themselves. This is more consistent with everyday experience (consider what attracts your attention). I prefer Freud's formulation: "consciousness arises *instead of* a memory-trace" (1920, p. 25; emphasis added). Perhaps the best way of putting it is that consciousness signals the need for new predictions; it signifies, as it were, "prediction-work in progress." Consciousness actually *changes* memory traces (cf. my brief discussion of reconsolidation in the Target Article).

I can easily understand why **Vittorio Gallese** considers my dichotomous classification of consciousness (affective-brainstem vs. cognitive-cortical) to be too rigid. He believes I claim that "phenomenal selfhood is the *exclusive* outcome of the upper-brainstem nuclei and of the limbic system" (Gallese). Actually, I agree with him that "a self whose experience of *encounters with the world* is constantly guided by the feelings such encounters evoke is inconceivable without the crucial role played by the neocortex" (Gallese). It is also true that "the body is primarily given to us as 'source' or 'power' for action—that is, as the variety of motor potentialities defining our interaction with the world we inhabit" (Gallese). However, speaking phylogenetically, it is important to remember that neural representation of the sensorimotor body at the tectal level long precedes the development of cortex. It is highly questionable whether this tectal (associative) body is capable of the cognitively conscious representational gymnastics that Gallese describes in the second half of his commentary. That consciousness of the affective consequences of the encounter with the world is *subsequently extended* to the body's cortical representations of the world, and is thereby stabilized, was never in doubt. Gallese is also unhappy with my characterization of cortex as random-access memory space. But how else are we to understand the findings of Mriganka Sur? Redirecting visual input from "visual" to "auditory" cortex reorganizes the latter to support completely competent vision (see my Target Article).

Jim Hopkins makes the same error that Freud made in 1923: he conflates the id with the system unconscious. Freud introduced his id concept to accommodate, among other things, the unconditioned and nonrepresentational nature of what he called "psychical energy." His "ego" concept theorized the manner in which the "id" (drives, instincts, and affects) are regulated, top-down, by memory traces (mental solids). Hopkins calls them internal objects. The system unconscious is derived from (hived off, excommunicated from) these regulatory processes (cf. Freud's "thing-presentations"). It is the system unconscious, therefore, that arises, as Hopkins puts it, in *consequence* of the "joint generation of motivation and consciousness," not the id. The structure of the id is innate by definition. If one were to similarly redirect an autonomic input to the brainstem, the result would be certain death—notwithstanding Heather Berlin's emphasis on neuroplasticity.

I have no difficulty with **Luba Kessler**'s suggestion that we should distinguish between subjective conscious states and "the self," and that we should probably reserve the latter term for stable (corticothalamic) *representations* of the subject. As she says, raw affective presence and the reflexive (or even the second-person) "self" are not synonymous. It is, however, important to realize that the self in her sense is an *object*-presentation. I like her poetic analogy, derived from the mirror-neuron concept, of the self "being akin to an internal mirror where the first reflection of the subjective state creates a "self" as the image of its own dawning representation" (L. Kessler). But mirror images are inconceivable without (cognitive) notions of spatial representation. Moreover, the states of the self described as "strong or feeble, cohesive or fragmenting, robust or disintegrating" (L. Kessler) involve quite complex forms of spatial cognition.

I find nothing that I disagree with in **Richard Kessler**'s commentary. But I would like to endorse his view that metapsychology is "the very hyphen in neuro-psychoanalysis!"

I cannot address all of **Larry Kunstadt**'s points here, so I will focus on the main ones. He, like Ariane Bazan, worries about my supposed conflation of id and affect. When at least two commentators are similarly misled, the problem must reside in the Target Article and not the commentators. For this I apologize and refer readers to my clarification above. Kunstadt also wonders: which claim of Freud's am I arguing against? The answer is: his claim that affective consciousness (like perception) arises from the superficies of the mental apparatus (Freud, 1923, p. 26). Next, Kunstadt complains that amending metapsychology on the basis of neuroscientific data conflates domains and "assumes we have solved the mind/body problem." In this respect, see the footnote in my reply to Bazan's commentary. As Richard Kessler says, metapsychology is the "missing link" between mind and body. Kunstadt goes on to assert that instincts are invisible and that "what counts to the analyst is what is visible." Drives and instincts are indeed invisible, but they are *experienced* as affects. Affects are no less real (no less observable) than objects, let alone words. Next, I want to address Kunstadt's important criticism that "if the [PAG] is lesioned and the lights go out, that does not mean that consciousness resides *in* the PAG" (emphasis added). I have partially addressed this point in my response to Heather Berlin. The essential issue, though, following the long-established neuropsychological principle of 'double dissociation," is that if psychological function *a* is lost with damage to brain region *A* but preserved with damage to region *B*, and psychological function *b* is lost with damage to brain region *B* but preserved with damage to region *A*, then one may legitimately claim that function *a* "resides in" region *A* as opposed to *B*, and that function *b* "resides in" region *B* as opposed to *A*. It is only in this strict sense that I want to assert that the capacity for consciousness as a whole "resides in" the upper brainstem as opposed to cortex. Actually, I prefer the term "arises from." Kunstadt's puzzlement about my alleged claim that the ego is "completely unconscious" (Kunstadt) is easily dealt with: I do not claim that it is completely unconscious, only that the consciousness of the ego is borrowed from the id and that the ego aspires to unconsciousness (to error-free predictions, automaticity). Similarly: of course I agree that "affect can modulate consciousness" (Kunstadt). To restate my basic argument: (1) affect *is* consciousness; (2) affective consciousness is not "blank"; (3) cognitive consciousness is bound affect, and the binding is always incomplete. Kunstadt's comparison of my theory with that of Barry Opatow (1997) is not on all fours: Opatow's is a theory of conscious and unconscious *representation*, not one of *affect*. Affect is always conscious. Next: I do not claim that "consciousness reads motor output" (Kunstadt). Even if Kunstadt here equates "motor" with all forms of free-energy discharge, I do not claim that. Consciousness attaches to bound energies, too. (The existence of conscious cognition is self-evident.)

Georg Northoff thinks I focus on *stimulus-induced* brain activity in relation to both affective and cognitive consciousness. This is not correct. I emphatically endorse the existence of *intrinsic* brain activity. However, I try to be more specific about which intrinsic activities are prerequisite for consciousness and which not. On the other hand, I am not persuaded that *content*, *form*, and *level* of consciousness are orthogonal parameters. Changes in level, for example, have immediate consequences for form and content. Northoff's further claim that I leave open the question as to "why there is consciousness at all" makes me worry that he misread my article. I am also surprised to hear that he considers this question to be the "hard" problem in current philosophy of mind. Chalmers's (2010) hard problem concerned the *how* question of consciousness, not the *why*.

Unlike Northoff, **Jaak Panksepp** has no difficulty recognizing that my theory of consciousness—based largely on his own—concerns "*intrinsic* dynamics of our subcortical neural networks" (emphasis added). See also his confirmation that my concern is more with the constitution than the correlates of consciousness. Unsurprisingly, Panksepp has no difficulty accepting that affective consciousness arises from the upper-brainstem and associated limbic structures. But does this imply that consciousness as a whole is constituted there? He asks: how does one negate the possibility that *cognitive* consciousness arises directly from cortex? "While anencephalic children and animals certainly provide compelling evidence for how much mind exists in subcortical domains, this does not unambiguously indicate that perceptual qualia emerged from affect" (Panksepp). This is where the double-dissociation paradigm comes into its own (see my reply to Kunstadt). Psychological function *a* (affective consciousness) is lost with damage to brain region *A* (upper brainstem) and preserved with damage to region *B* (posterior cortex), but psychological function *b* (cognitive consciousness) is lost with damage to either brain regions, *B* or *A*. Thus, one may conclude that function *b* does not "reside in" region *B*. It resides in a combination of regions *A* and *B*.

Manos Tsakiris advances a similar argument to that of Luba Kessler and some of the other commentators: he says that ego and id *co-constitute* self-specificity. He asks: "Is memory space the sole, or at least the most important, contribution that the ego can make to the id?" My answer is "yes," so long as one recognizes that the various cognitive functions that the ego performs (such as those that Tsakiris discusses) are all grounded in or derived from this basic representational capacity. He goes on to question my view that external body "ownership" is just one such acquired representation, albeit the most fundamental one. He quotes William James—"the same old body always there"—and reminds us that "my *real* external body . . . never leaves me" (Tsakiris). The "real" is not so simple in psychology. Psychoanalysis studies first-person perspectives on reality, not third-person ones; this is what distinguishes it from other sciences. Each human subject constructs a reality of his or her own. From this perspective, it cannot be said that *my representation* of my body "never leaves me." Consider, for example, out-of-body, autoscopic, and *Doppelgänger* experiences, as well as ideas of reference, etc., not to mention the body-swap and rubber-hand illusions that Tsakiris himself addresses. However, please note that I am not claiming on this basis that "the external body is an illusion" (Tsakiris). The more my conception of the relationship between me and my body approaches yours—that is, the more it approaches the third-person perspective—the less "illusory" it becomes (and the less prone to prediction-error, to free energy, to affect). But it will always remain a conception, a representation, an "idea." Tsakiris makes an interesting point when he argues that pre-representational external sensory states are no less primary and subjective than internal affective ones. I agree with him; however, at that level of processing (before the states in question congeal or coalesce into mental solids) they are precisely *not* representations, they are external influences on affect. (Cf. my response to Gallese, who makes the same claim for pre-representational *motor* states.) A similar point applies to Tsakiris's remarks about heartbeat. Heartbeat is a representation, a mental solid. It is quite different from the *affective* state of, say, anxiety, which one might feel alongside a racing heartbeat (see the footnote in my reply to Bazan). Again: I do not subscribe to the James–Lange theory. Affect is not the same as the interoceptive aspects of the external sensory modality for somatic sensation.

What **Oliver Turnbull** reconstructs about my own thought processes seems perfectly plausible. I agree that neuropsychoanalysis is faced with a major disagreement concerning the brain basis of subjectivity, in which the classical views of Craig (etc.) are contrasted with those of Panksepp and Damasio. In summing up their respective arguments at our Berlin Congress, I had to align the contrasting views with those of Freud. In doing so, it was obvious that, although the available evidence best supports the conclusions of Panksepp (etc.), Freud's conception coincided with Craig's. I have little doubt that Freud himself would have changed his mind when confronted with the facts now available to us.

Gerald Wiest appears to accept my main arguments, but he asks: "how can the id as a hallucinatory, unrealistic, not time-bound, and irrational agency by nature represent a conscious entity?" In other words: how can all of this be conscious simultaneously with the veridical, realistic, time-bound, and rational consciousness of my ego? My answer is that it is not. The *representational* attributes that Wiest enumerates are attributes of "the unconscious," not of "the id" (see my response to Hopkins). The id's contribution to consciousness is the affect that accompanies the ego's representations. The ego's contribution is to convert the affect into ideas, which are conscious ideas in the first place, and then (by progressively improving the realism—the predictive power—of the ideas) to render them unconscious. In other words, the ego's contribution is firstly to minimize affect and ultimately to minimize consciousness as a whole (to minimize free-energy "surprise," in Friston's terms). If the ego renders an idea unconscious before it has mastered the bit of reality it represents, then that idea will possess the unrealistic, etc., attributes that Wiest enumerates. But it will still be a piece of ego, albeit a piece that is excommunicated from ego consciousness ("the repressed").

Last but not least, a few remarks about **Maggie Zellner**'s largely clinical commentary. First, when she speaks of "devoting disproportionate resources to representing the other's image of us, rather than maintaining our own perspective," she is of course referring to common-or-garden superego pathology ("the other's image of us" is what the superego is made of). It is best to ground new theories in the old ones. Second, when she discusses the mutative mechanisms of psychoanalytic treatment, she does not sufficiently emphasize reconsolidation, which I think is quite a fundamental mechanism of this type. Finally, when she says that she does not advocate "living purely 'in the id'," I hope it is clear that nor do I. An ego-less id—unless cared for by others (e.g., parents)—cannot begin to cope with the world. It is therefore doomed to a life overwhelmed by affect (constant "surprises") followed rapidly by certain death.

In closing, I would like to express my appreciation to all the commentators, who expended so much time and effort on my ideas. I am grateful to them for the opportunity to clarify what I meant and did not mean. But more importantly, I am grateful for the demands their comments made on me to clarify my own thinking, for myself. I have seldom confronted a more difficult intellectual task than the one addressed in this article; I therefore need all the help I can get!

REFERENCES

Chalmers, D. J. (2010). *The Character of Consciousness.* Oxford: Oxford University Press.

Freud, S. (1915a.) Instincts and their vicissitudes. *Standard Edition*, 14.

Freud, S. (1916–17). *Introductory Lectures in Psychoanalysis. Standard Edition*, 16.

Freud, S. (1923). *The Ego and the Id. Standard Edition*, 19.

Freud, S. (1926). *Inhibitions, Symptoms and Anxiety. Standard Edition,* 20.

Freud, S. (1940). *An Outline of Psychoanalysis. Standard Edition,* 23.

Opatow, B. (1997). The real unconscious: Psychoanalysis as a theory of consciousness. *Journal of the American Psychoanalytic Association*, 45: 865–890.

Parvizi, J., & Damasio, A. (2003). Neuroanatomical correlates of brainstem coma. *Brain*, 126: 1524–1536.

Solms, M. (1997). What is consciousness? *Journal of the American Psychoanalytic Association,* 45: 681–778.

Solms, M., & Turnbull, O. (2002). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. New York: Other Press.