Affective Neuroscience and the Philosophy of Self

Mammalian Agency and The Self: Implications of Affective Neuroscience

Stephen T. Asma

Stephen T. Asma
Professor of Philosophy
Fellow at the Research Group in Mind, Science and Culture
Columbia College Chicago
Chicago IL 60605 USA
Email: sasma@colum.edu

The nature of self awareness and the origin and persistence of personal identity still loom large in contemporary philosophy of mind. Many philosophers have been wooed by the computational approach to consciousness, and they attempt to find the self amidst the phenomenon of neo-cortical information processing. Affective neuroscience offers another pathway to understanding the evolution and nature of self. This paper explores how affective neuroscience acts as a positive game-changer in the philosophical pursuit of self. In particular, I will focus on connecting "mammalian agency" to (a) subjective awareness, and (b) identity through time.

I. The Problem of the Self

What am I? I am obviously an individual person, observable by others –a public physical organism. I can be picked out of a crowd. But to myself, I am a subject; an agent moving through the world with a rich inner life of thoughts, feelings and memories. I am a self.

In its modern formulation, the philosophical problem of the self goes back to Descartes and David Hume. But the puzzles of self-identity are perennial (maybe even inevitable) and stretch back to the Ancient Greeks and the Vedic and Upanisad literature of the Hindus. The ancient Greek playwright Epikarmos even tells a story of a man who borrows money from his neighbor, but when he is pressed to repay the loan, he reminds the courts that he is, like every other natural thing, constantly changing (--what with the ceaseless exchange of matter) and can't literally be said to be the *same* guy who borrowed the money awhile back.

All expedient philosophy aside, self-identity has been a longstanding puzzle. How does a subjective unity emerge out of a plurality of mental abilities? How does self-reflective awareness relate to those abilities? And how does self-identity persist (with continuity and change) over time? Jaak Panksepp's affective neuroscience brings fresh perspectives to the philosophy of self. In order to appreciate these fresh perspectives, we need to situate ourselves a bit in the modern conversation.

David Hume pointed out Descartes' error (but not the "error" that Antonio Damasio focused on).[1] According to Hume, Descartes had no right to think of the "I" as a metaphysical substance. The *cogito ergo sum* does not establish the existence of metaphysical substance –it only proves the existence of momentary self in each act of thinking. But now Hume found himself in a new dilemma. If all ideas –all knowledge – originates in sense impressions (a basic Empiricist commitment), then what should we make of the self? My self cannot be found as a discrete content of consciousness –it is always the knower and never the known.[2] Hume concluded counter-intuitively that I am really just a bundle of experiences (memories, emotions, cognitions, etc.) and the self is a kind of fiction.[3] Following Hume, Kant continued a more *functional* approach to the self, rather than a naïve metaphysical view. The self is the point of unity or focus of subjective perception, feeling, cognition –but the self must be presupposed or inferred in order to make sense of experience. The self is not a fiction, but it is also neither directly experienced (through the categories of understanding) nor directly encountered through intellectual intuition.

Many contemporary philosophers have continued this tradition. The self accompanies the content of experience with something like an "awareness tone" --and this moment of self-awareness, this crystallization of subjectivity, is a "thin subject" lacking "ontic depth" (Strawson, 2009). This very rarefied high-level self is also exceedingly promiscuous. It flits about and colors whatever experience is currently underway. This translucent self is a movable awareness that emerges in different

---

[1] Antonio Damasio's book Descartes' Error: Emotion, Reason and the Human Brain (Penguin, 2005) famously argues that Descartes' great error was thinking of the mind as separate from the body (and therefore, the emotions). Descartes' dualism is more complicated however, and Damasio's critique takes a somewhat uncharitable view of Cartesian mind. It is true that mind and body are metaphysically distinct, according to Descartes, but he never viewed mind as a purely rational calculator detached from emotional life. His point was that bodily affects are not a part of the subjective life until they can be read-out (as emotions or feelings) by the conscious mind. Descartes' contemporaries and next-generation philosophers like Hume, however, saw a different "error." For Hume and Kant the mistake was thinking that a conscious unity of experience (the cogito) proves the existence of a corresponding entity –an ontological self. No positive metaphysics can be derived legitimately from the cogito.

[2] Hume says: "For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception of other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of myself, and may truly be said to not exist." (Treatise I, iv, 6,)

[3] Interestingly, the Buddha makes similar arguments in the *Potthapada sutta* (DN) against the metaphysical notions of *atman* and also against the notion of a separable consciousness –a res cogitans. In the *Mahatanhasankhaya sutta* (MN), he likens consciousness to fire, and fire exists only on the fuel it burns –never in some pure disembodied form.

functional modes, but has no personality per se.[4] Where is my real self, for example, when I'm struggling with a Boolean algebra problem? In this case the self seems to "reside" in the higher neocortical activities of mathematical thinking, but if you suddenly poke me with a pointed stick, then my self will quickly shift to the material body domain. Each new activity –indeed each new moment --brings a new self. If there is such a diaphanous self, then not much can be said about it at this point. One wonders, however, whether we may one day marry the phenomenological self-report of the self-aware subject with sophisticated brain imaging in a way that reveals some unique recursive neural reverberation. We may one day find some neural flash that serves as the material substrate for our familiar sense of translucent subjectivity. This subjectivity is probably an emergent property of various neurochemical systems, some of which reach way down into the limbic and possibly subcortical levels.

Below this arid domain of the philosopher's translucent self, however, lies the realm of self that most laypeople contemplate. Here is the self of common sense. A self that has personality –built up over time with beliefs, memories, and life history. William James and pragmatists like George Herbert Mead reminded philosophers that subjectivity is not utterly pure, but mixed and integrated with social life.[5]

Philosopher Daniel Dennett describes this more content-rich self as our "center of narrative gravity" (Dennett, 1988). Antonio Damasio calls this our "autobiographical self" (Damasio, 2000). And as these names suggest, this self is largely composed in the highly discursive process of neocortical reflection. Hubert Herman's psychological theory of the dialogical self draws heavily on this tradition (Herman and Kempen, 1993).

Language, together with frontal-lobe powers, allows us myriad ways to represent the world and represent ourselves. We make ourselves, at this level, through the stories we tell ourselves. Many of those representational processes (that govern our self identity) will be constrained by those rules of cognition that computational cognitive science seeks to isolate. And all the relativism not withstanding, the social constructionists have also recently helped us to better appreciate the role that society can play in this narration of

---

[4] My use of the term "translucent self" is perhaps idiosyncratic, but it responds to a contemporary discussion in the phenomenology of self. Phenomenologists like Thomas Metzinger (2003) and Dan Zahavi (2005) have developed the terminology of "transparency" and "opacity" of the self in rather precise ways. Metzinger, for example, describes the phenomenologically transparent self as a way of describing a pre-reflective state of naïve experience (being in a world), wherein the representational (and perhaps agency) aspects are *invisible* to the subject (i.e., the self is a transparent "window" through which the subject sees the contents of experience, and only the contents are attended to). The phenomenologically opaque self is when I am aware of my own representational processing –I attend to myself as the "vehicle" or the "framer" of the content of experience, as in the case of pseudo-hallucinations or lucid dreaming. My own use of the term "translucent self" is partly to acknowledge this interesting discussion in the literature, but also to reject the dichotomous tendency of this form/content distinction. One of the implications of the "mammalian agency" approach to the self may be that subjectivity is never purely opaque nor transparent, but somewhere in between.

[5] William James offers a compelling integration of self theories in Chapter X "The Consciousness of Self" (James 2007).

self identity. But while all of this is fascinating and while good work will continue at this level, Dr. Panksepp's revolutionary work wants to take us lower still –into the ancient, unexplored but powerful sources of self.

II. Mammalian Agency

In contrast to the neocortical, highly linguistic aspects of mind, Dr. Panksepp goes down to the foundations of mammal agency. In doing so, he develops a more capacious concept of consciousness –one that includes emotions and their primitive affects –and expands our notion of mind beyond the representational and propositional versions that dominate both cognitive science and traditional philosophy. Affective neuroscience reminds us of the *body* and its *non-linguistic* forms of *meaning.* Dr. Panksepp revises Descartes' cogito, claiming instead "I feel, therefore I am" (Panksepp, 1998). But even deeper than this limbic consciousness he pursues the primitive SELF (Simple Ego-type Life Form) in the prelinguistic motor-mapping system of the ancient midbrain.

Dr. Panksepp's archaic self is a biological notion of identity. It is a concept of self based more on *action* than rarefied intellectual *reflection*, and so it includes many other kinds of nonhuman animals in the club of selves. An organism trying to evade a predator, within a specific environment, is solving a multitude of challenges in real time. It does so from a specific point of view in space and time –constantly adjusting its body and modulating behaviors. A rabbit trying to evade a predator, to use Dr. Panksepp's example has little conscious sense of its own future and past (given the reality of its modest frontal lobes) but "It is dealing with its present circumstances on a moment to moment basis. It is precisely those here-and-now states of consciousness that we must seek to understand before we can grasp how they come to be extended in time, as they are in the human mind through our frontal cortical time-extending and planning abilities"(Panksepp, 1998).

Affective neuroscience reminds us of our phylogenetic homologies with other mammals, and so our biological identity should be found near the core of the brain --not the more recent neocortex. This archaic SELF would be a basic motor-mapping system – a template for action tendencies. Despite the inclination of philosophers to think about consciousness and subjectivity in terms of perceptions (like sense data qualia), affective neuroscience reminds us that "a level of motor coherence had to exist before there would be utility for sensory guidance." (Panksepp 1998) This archaic SELF would have to coordinate or integrate emotions from the periaqueductal gray (PAG) region of the brain and the perceptual somatosensory system. The centromedial zones of the brain (especially the deep layers of the colliculi and the PAG) answer to this requirement. Moreover, Dr. Panksepp's experimental work with mammals suggests that this area is much more relevant to biological intentional identity than higher neocortical areas. Experimentally induced lesions along the PAG are much more devastating to the intentionality or seeming *agency* of the animal than lesions in the higher areas of the brain. This archaic level of self is not cognitive. It is what Dr. Panksepp calls "primary process consciousness" and it resides in the intrinsic action-readiness of the biological system.

Beyond the simple integrated motor actions of this SELF, it is also likely that this centromedial zone provides a "coherent matrix in which a variety of sensory stimuli become hedonically valenced." (Panksepp 1998) In other words, the organism is establishing attraction and aversion values at the subcortical level, and so the organism's

most rudimentary self-awareness, of a spatio-temporally located body in an environment, will already be coded with positive and negative affects.[6] The self is not superadded after a certain level of cognitive sophistication is achieved (a view commonly held by philosophers). Rather, the self first emerges in the precognitive ability of most organisms to operate from an egocentric point of view. Way below the level of propositional beliefs, animals must solve basic motor challenges (e.g., where am I in relation to that advancing sharp claw thing? Am I moving now, or is the environment moving? Am I eating my own arm?). For mammals this low level ability is accompanied by the archetypical survival systems, shaped by natural selection over geological time. These are the homological affective systems that Panksepp has isolated in the brains and behaviors of his test subjects: approach when SEEKING, escape from FEAR, attack in RAGE mode, pursue nurturance in PANIC, seek mate in LUST mode, and so on. These affects and emotions are survival skills and comprise primary and secondary consciousness –they have to be "owned" by the organism for them to work properly. This is why Panksepp and Damasio, both fans of Spinoza's monism, are in agreement about the reality of primary or core consciousness.[7] Subjectivity resides first in the biological realm of action. It is not the disembodied Cartesian spectator.

III. Philosophical Implications

Now what are some of the implications of this notion of self? First, like other forms of scientific naturalism, it demonstrates that we do not need additional metaphysical agencies (like souls, or noumenal mental realms, etc.) in order to explain personal identity or even subjectivity. Secondly, and more significantly, Panksepp's archaic self – with its primary consciousness –rescues the body and feelings from the long philosophical tradition that characterized them as purely unconscious machinery. We all know of Descartes' dualism-derived "animal machines," but even David Chalmers (1997) seems to think that fully functioning animals with intact brains and bodies could be zombies –"all is dark inside" with nobody home. Panksepp's approach suggests that consciousness is not superadded to otherwise functioning survival machines, nor can consciousness be abstracted out of the physio-chemical system (except I suppose in parlor-game thought experiments). Even Dan Dennett, who is usually quite sensitive to the biological sciences, offers an example that betrays a cognitive bias about consciousness. He asks, "What is it like to notice, while sound asleep, that your left arm has become twisted into a position in which it is putting undue strain on your left

[6] Panksepp (2005) suggests "there exists a subcortical viscero-somatic homunculus, laid out in motor-action coordinates, that creates a primal representation of the body (core SELF) that can be modulated by global brain emotional networks that establish affective intentions in action, which are projected onto the world as prototypical affective values, helping guide cognitive intentionality."

[7] To understand what it means to have raw affective feelings, Dr. Panksepp suggests that "we must entertain neuro-psychological conceptions of human and animal 'souls' through concepts such as the 'core self' (Damasio 1999; Galagher & Sheard, 1999; Panksepp, 1998a). I suspect our mental lives are critically linked to primal viscero-somatic representations of the body situated in paramedian regions of the brain, and connected to associated higher limbic areas…." (Panksepp, 1998).

shoulder? Like nothing: it is not part of your experience. You swiftly and unconsciously shift to a more 'comfortable' position…." (Dennett, 1996). And Dennett concludes that whatever "clever" problem-solving is going on at these biological levels, it is not a part of our mental lives at all. Of course, Dennett and others have a point here. Many of our brain-based competencies (like the autonomic systems) happen below the radar –but Panksepp's approach offers the tantalizing possibility that we can get into the muddy unconscious. In fact, with his notion of primary subcortical consciousness, he seems to be changing the game and eliminating the traditional notion of an unconscious.[8] After all, in Dennett's own example of the sleeping subject, I do not move my arm in *any* chaotic manner –I don't fling it into my face, or put it into a less comfortable position. My primal self solves the problem with a somewhat nuanced sense of the spatio-temporal environment and the relevant motor possibilities. We can say, as Dennett does, that this has no connection with our mental lives at all, but this only betrays an overly narrow conception of mind (e.g., neocortical computation, or what Panksepp calls tertiary consciousness). We may not have much first-person phenomenological data of this archaic self or this mind, but significant access can be gained by the kinds of experimental brain manipulations that guide affective neuroscience research.

One of the most interesting implications of this biological notion of self-identity is that it answers some of the traditional skepticism about the self. From the Buddha's criticisms of atman, through Hume's bundle theory, and up to today's postmodern rejection of an essential core, these skeptical traditions have adopted the decentered subject. But if Panksepp is right, then the fracturing of the subject is overestimated, and the embrace of a decentered self is premature. Yes, the diaphanous self is momentary and cannot be directly observed inside experience, but the so-called "binding problem" of apperception may be more imaginary than real. Panksepp's SELF gives us a way in which the fleeting and ontically thin "I" keeps getting referred back to the biological "I." The centromedial zones of the mind-brain produce a primitive self that *persists over time*, because it is a "central processor" of inputs and outputs for an organism that *is* extended in space and time. The fleeting I of the cogito may be reborn during every change in emotional or perceptual or cognitive content, but much of the content of our experience (perhaps all of it) will first be organized by the centromedial zones of the midbrain and the core affect systems. So, one of the implications of Panksepp's work is showing how the higher rarefied subjectivities of self may find constant tether to our very specific animal identity. Hume, who didn't have the benefit of living after Darwin, went looking for the self in the wrong part of the psyche –namely in the representational mind.[9]

---

[8] "The traditional answer has been that one does not have any mental experiences until certain kinds of information interact with --are 'read out' by –higher neo cortical mechanisms that elaborate our awareness of the world. Many still believe that affects are not experienced in the lower reaches of the brain –that all brain functions below the neocortex are experientially implicit and unconscious. Within such anthropocentric world-views, emotional feelings cannot be understood until we figure out how the higher regions of the brain generate awareness of the world." (Panksepp, 2005)

[9] The irony is that Hume uniquely grasped the overwhelmingly passional/emotional nature of human beings –demoting reason down to lowly "rationalizer" rather than imperious controller. But then, prior to Darwin, Hume had no real way to connect

Strangely enough, contemporary philosophers are still looking in the wrong, albeit well-lit place. Panksepp, however, is finally delivering on the Darwinian promissory note: a subject, an "I", that is truly born out of the struggle for survival. The binding problem is not a problem because subjectivity is always content-laden with the unified life of the spatio-temporally located organism and its evolved archetypical dispositions. In this way, I think Panksepp's and Damasio's solutions are somewhat similar, though Panksepp seems more explicit in locating such identity in the ancient brain.

Damasio's most recent book, *Self Comes to Mind*, attempts to clarify the similarities and differences with Panksepp's long-held theory of a primary self. Circumspect about his own work, Damasio explains that his previous accounts of the self were focused too high up in brain processing, and now he recognizes a brainstem based "primordial" or "proto self." This proto self, according to Damasio, corresponds more with Panksepp's primary SELF, but Damasio wants to locate it even lower down the brainstem (nucleus tractus solitarius) than Panksepp suggested (periaqueductal gray). "In the perspective of evolution" Damasio says, "and in the perspective of one's life history, the knower came in steps: the protoself and its primordial feelings; the action-driven core self; and finally the autobiographical self, which incorporates social and spiritual dimensions." (pg. 10)

Panksepp and now increasingly Damasio want to locate a fundamental self deep in the real-time processing of mammal brains, but what remains contentious and empirically unverified is whether that locus is more in the motor structures (Panksepp) or in the sensory structures (Damasio).[10] It's hard to see how this disagreement will be resolved. Brainstem processes produce felt body states in the organism, and these primitive sensations of pain and pleasure are intimately integrated with the action-orientation of the motor systems.

The points of agreement between Damasio and Panksepp are many however, and perhaps the most important is the way both affective scientists marshal impressive data to demonstrate that the self is not just a product of the cortex. Panksepp's work with decortication of rats is well known,[11] but in *Self Comes to Mind* Damasio strengthens the argument significantly by taking us into the rich emotional life (grounded in the proto self) of children born without a functioning cerebral cortex. Damasio argues that these children demonstrate low-level agency and basic levels of emotional integration.

Damasio's brand of affective science, which also tries to get all the way up into the higher levels of cognitive life may have more to offer philosophers who are interested in the uniquely complex subjectivity of human mind. Self identity over time is woven together, according to Damasio, in the "autobiographical self" which at first sounds like

---

(logically, let alone chronologically) the limbic life with the rational. Subsequently, the affects slowly submerged into a swamp of philosophical incognita.

[10] See footnote 17 of Chapter One (2010) for Damasio's clearest articulation of his difference with Panksepp regarding self.

[11] "Effects of neonatal decortication on the social play of juvenile rats" by Jaak Panksepp, Larry Normansell, James F. Cox, and Stephen M. Sivly in *Physiology and Behavior* Vol. 56, Issue 3, 1994.

the discursive representational narrations of higher neocortical processes.[12] There's no doubt that big-brained *Homo sapiens* can spin elaborate coherence out of disparate experience, using memory, discursive rationality, and intentional projections. But combinations of non-linguistic perceptions, like visually based image schemas, together with engraved feeling dispositions may be all that is necessary to begin some rudimentary *autobiography* of self. Animals with very impoverished symbolic and conceptual skills may nonetheless have the ability to sense (literally) their own personal history and then comport themselves into the near future (again, drawing on their affective entrenchments, rather than cognitive reflections).[13] Nonetheless, many philosophers are more captivated with the truly symbolic manipulations of the human autobiographical self. For these philosophers, Damasio will be more intriguing than Panksepp.

Ultimately, skeptics about the self have been right to scoff at the idea of a mysterious transcendental homunculus that sits like a spectator in a Cartesian theater. But they were wrong to dispense with agency. It is often said of bundle theorists, whether the Buddha or David Hume, that they want to characterize *thinking* without a *thinker* –or they want to get the thoughts to *think themselves*. These are laudable moves as philosophers try to account for the invisibility of the self, but perhaps these counterintuitive moves are the unfortunate product of doing one's philosophy in the neocortical paradigm of representation and perception. Go lower into the biological agency of affective consciousness and the idea of a self that collects, unifies, and weights content (but remains invisible to tertiary consciousness) makes more sense.

---

[12] In the first chapter of Damasio's The Feeling of What Happens (Mariner Books, 2000) he seems to suggest that the promiscuous self (the ontically thin "I") is something that accompanies the real time here-and-now experiences of many non-human animals. But these core-self subjects must be woven together into a coherent record of the organism's life history. For humans this weaving will be heavily cognitive, volitional, and reflective, but for other mammals it will be more deterministic neurological engraving. Damasio recognizes, in Chapter Six, that autobiographical composition of the self can be non-linguistic and image based. His latest work further strengthens the idea that certain kinds of self are pre-linguistic.

[13] It's my view that this might be an interesting meeting place between affective neuroscience and the metaphor-based epistemology of philosophers like Mark Johnson (2007) and George Lakoff (1980). Representational cognition is obviously very sophisticated when compared with sensual problem solving in lower animals, but I doubt that it emerged as a *sui generis*. The progenitor of propositional conceptual knowledge must be bodily knowledge, which in turn must be more image-based, affect-based, and spatially, temporally relative to our particular evolution. The metaphorical root (what Johnson calls "aesthetic" and I would call "affective") of cognition, is just one more reason why the computational model of mind is unsatisfactory.

References

Buddha "Potthapada sutta" *The Long Discourses of the Buddha: A Translation of the Digha Nikaya* (trans. Maurice Walshe) (Wisdom Publications, 2nd ed. 1995).

Buddha "Mahatanhasankhaya sutta" *The Middle Length Discourses of the Buddha: A Translation of the Majjhima Nikaya* (trans. Bhikkhu Bodhi, Bhikkhu Nanamoli) (Wisdom Publications, 2nd ed. 1995).

Chalmers, D. (1997) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press).

Damasio, A. (2000). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Mariner Books).

Damasio A. (2005) *Descartes' Error: Emotion, Reason and the Human Brain* (Penguin).

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

Dennett, D. (1988). "Why We Are All Novelists" *Times Literary Supplement* (Sept. 16-22, 1988)

Dennett, D. (1996). *Kinds of Minds* (Basic Books).

Herman, H. and Kempen, H. (1993). *The Dialogical Self: Meaning as Movement* (Academic Press).

Hume, D. (2009) *A Treatise of Human Nature* (Merchant Books).

James, W. (2007). *The Principles of Psychology Vol. I* (Cosimo Classics)

Johnson, Mark (2007). *The Meaning of the Body* (University of Chicago Press).

Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By* (University of Chicago).

Metzinger, T. (2003) “Phenomenal Transparency and Cognitive Self-Reference” *Phenomenology and the Cognitive Sciences* 2: pp. 353-393.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions* (Oxford University Press).

Panksepp, J. (2005) “On the Embodied Neural Nature of Core Affects” *Journal of Consciousness Studies* 12, No. 8-10, 2005, pp. 158-184.

Strawson, G. (2009) *Selves: an Essay in Revisionary Metaphysics* (Oxford, 2009).

Zahavi, D. (2005) *Subjectivity and Selfhood* (MIT Press).