

A Review Essay on Antonio Damasio's *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*

Aldo Mosca

66 West 12th Street, Room 604

New School University

New York, NY 10011

U.S.A.

moscaa@newschool.edu

<http://homepage.newsouth.edu/~moscaa>

Copyright (c) Aldo Mosca 2000

PSYCHE, 6(10), October 2000

<http://psyche.cs.monash.edu.au/v6/psyche-6-10-mosca.html>

KEYWORDS: emotions, consciousness, self, neuropathology.

REVIEW OF: Antonio Damasio (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace. 386pp. ISBN: 0156010755. US\$15 pbk

Abstract. Here I present a sympathetic but critical analysis of Damasio's latest book. I begin with a brief exposition of his neurobiological theory of emotion and then take issue with him on the cognitive and representational nature of emotional states. I discuss his view of consciousness as a second-order set of neural activations, which are allegedly intimately related to, and even necessary for, elementary emotional and homeodynamic processes. I find Damasio's account wanting in at least two respects. First, the relationship between emotional states and the consciousness thereof is left ambiguous. Second, Damasio lacks a clear psychological hypothesis about the hierarchy of first-, second-, and possibly third-order mental/neural states. The result is that a wealth of neurobiological information fails to be organized in a coherent conceptual scheme, and the whole account is ultimately unsatisfactory.

1. Introduction

Antonio Damasio and his colleagues have been working for several years on the fundamental assumption that neuroscience has mostly explored strictly cognitive processes, whereby the brain gathers and manipulates information about the environment, but has overlooked the important relation between the brain and the body or soma, including emotional processes. This is an unfamiliar and often neglected sort of physiological dualism distinct from the classic mind/body problem, yet Damasio also attributed it to Descartes in his 1994 book, *Descartes' Error* (Damasio, 1994). The culminating point of that work, among many insightful and fascinating suggestions, was the "somatic marker hypothesis," according to which -- very briefly -- rational decision-making would be in fact conditioned by observable emotional somatic responses that a subject uses as clues to the "good" or the "bad" of a given prospect. Classical Cartesian dualism was explicitly ruled out, but a mentalistic vocabulary was preserved on the grounds that neuroscience is not yet in a position to offer a complete account of the relation between the physical and the mental.

In his new book, *The Feeling of What Happens*, Damasio tackles the problem of consciousness already outlined towards the end of his previous work, and he characteristically sets his discussion against the background of a neurobiological theory of emotions and feelings rooted in the body. Specifically, two problems of consciousness are identified. The first is to explain how neural patterns engender mental "images", where an image is not just a visual percept but any mental pattern built with the tokens of each sensory modality. Damasio acknowledges an "explanatory gap" (see Levine, 1983) between neurobiology and the "simple sensory qualities" of experience, but is confident that scientific research will bridge it. The second problem of consciousness is to explain how we come to have "a sense of self in the act of knowing" (p. 9). Damasio is concerned almost exclusively with the latter problem -- "There is no consciousness that is not self-consciousness" (p. 20) -- and especially with its relation to the knowledge or feeling of emotions.

The central thesis of the book is that self-consciousness -- "core consciousness" in Damasio's terminology -- is a second-order state of the mind/brain located in some specific regions, that is capable of representing the relation between representations of objects and representations of the soma, while the latter is almost invariably reacting emotionally to some object or another. (In Damasio's terminology, "object" may refer to a thing external to the body but also to a toothache or a state of bliss; this may be the source of some unnecessary confusion, which I will try and avoid). The thesis is an elaboration on his former claim that feeling is "the realization of a nexus between an object and an emotional body state" (Damasio, 1994, p. 132). Damasio further distinguishes core consciousness, which is concerned with the here and now, from extended consciousness, which includes autobiographical memory and the perception of time. His method of inquiry, as always, is to advance theoretical hypotheses as suggested and supported by clinical and anatomical evidence.

The theory of emotion presented in the text is not new, except for a and problematic distinction between feeling and consciousness, about which more below. I will first present a synopsis and then offer some comments. This will set the stage for my discussion of Damasio's theory of consciousness.

2. A Theory of Emotion.

The emotions are defined as patterns of chemical and neural responses, the function of which is to assist the organism in maintaining life by prompting adaptive behaviors. They are due to the activation of a set of brain structures, most of which also monitor and regulate bodily states around optimal physiological values, in processes known as homeostasis or homeodynamics. The emotions are biologically determined, stereotypical, and automatic, although it is acknowledged that culture and individual development may influence the set of inducers and can inhibit or modify overt expressions.

Damasio distinguishes: (a) six primary or universal emotions, namely happiness, sadness, fear, anger, surprise, and disgust (the list is the same as in Ekman, 1992, and is based on universal facial expressions); (b) other "behaviors" or secondary and social emotions, like embarrassment, jealousy, guilt and pride; and (c) a class of background emotions such as well-being or malaise, calm or tension, fatigue or energy, anticipation or dread. In the background emotions, the inducer is normally internal and the focus of response is mainly the "internal milieu" of the body.

There are rather well defined brain regions responsible for emotion. The brain stem (at the top of the spinal cord) is involved in virtually all of them; the hypothalamus (a subcortical structure) and the ventromedial prefrontal cortex are responsible for sadness, but never for anger or fear; and the amygdala (another subcortical structure) is mainly responsible for fear and for the recognition of fearful expression. The hypothalamus and brain stem locations are presented as recent and novel results in brain research (but see also Damasio, 1994, and LeDoux, 1996). The anterior cingulate cortex (a "belt" surrounding the corpus callosum, which connects the two cerebral hemispheres) and the basal forebrain are also mentioned; this will turn out to be somewhat problematic, because the cingulate cortex is regarded by Damasio as the most important region among those which make consciousness possible, and it is by no means clear whether consciousness is necessary for emotion. In any event, this is a somewhat broader view than that presented in Damasio (1994), where the amygdala and the cingulate cortex were prominent.

The physiology of emotion is not easily described, because of the simultaneous activation of several structures and feedback reactions among them. This prevents the process from being straightforwardly linear and creates significant problems for an overall conceptualization of levels or orders of processing. Damasio does his best, and so will we.

1) An object or situation is either detected and represented by sensory devices or recalled from memory, and these neural patterns activate nuclei in the brain stem, the hypothalamus, and the amygdala. This perhaps is an emotional episode, but it can also be regarded as the first step in a process whose final consequences are, for example, heart-rate alterations, facial expressions, and characteristic behavior. Whether the ventromedial parietal cortex, the basal forebrain, and the anterior cingulate cortex are automatically activated from the beginning, or whether they have an intermediate role in the process, not necessary for elementary activation, is not exactly clear. According to LeDoux (1996), for example, the amygdala can be activated without the intervention of cortical areas, although for a period no longer than twelve milliseconds (Damasio has no comment on this).

2a) The structures just mentioned release hormones of various kinds in the bloodstream, directed both towards the body proper, thus changing the chemical profile of the internal milieu, and towards other brain regions, such as the somatosensory cortices and the cingulate cortex, thus modifying the signaling of body states to the brain. Note that the activation of the cingulate cortex appears here as a second step in the process.

2b) The same structures simultaneously send faster electrochemical signals via neurotransmitters, both to bodily organs such as the adrenal gland, which releases hormones, which in turn reach back to the brain, and directly to other brain regions such as the cerebral cortex, the thalamus, and the basal ganglia, thus modifying what Damasio calls the "cognitive state", namely the propensity to exhibit some emotional behaviors and the speed and focus of mental processing.

A warning: toward the end of each of the two last paragraphs we may have already reached the level of "feeling", which for Damasio is the experience of an emotion, or the sensing of the consequences of emotional activation. The localization of feeling, its distinction from emotion, and its relation to core consciousness will turn out to be significant difficulties in Damasio's whole account.

Some comments. First of all, Damasio's account of the emotions is unabashedly non-cognitivist, since it describes an emotional process as a set of cerebral, somatic, and behavioral responses to the perception or memory of an object. Memory drops out of consideration after the process's inception and is not construed as built into the emotional state. (The representation of the object, it must be said, will reappear in the following, just as it did in *Descartes' Error*, but is certainly not prominent in Damasio's new introductory discussion. Incidentally, the fact of the matter is that Descartes too denied a representational nature to emotional states and regarded them as the mind's sensings of brain events, the latter being caused in turn by sensory stimulations. See Descartes, 1649, article 27.) Now this approach to the emotions happens to be at odds with the results of thirty-odd years of research in cognitive psychology (see, e.g., Frijda, 1986; Lazarus, 1991) and is rather closer to the position of William James (1884), which Damasio not surprisingly still defends. Unlike James, who was utterly ignorant of special emotional processes in the brain,

Damasio of course knows them very well, yet he still confines the object's representation to the role of a stimulus and nothing else.

There are several problems with this account, but perhaps the greatest difficulty lies in its incapacity to situate an emotional episode in the general economy of the mind/brain. What remains unexplained is how a subject in a state of autonomic arousal, visceral contractions, and an increased heart-rate goes about devising a solution to her or his problem without relying on information about the nature of the relevant object, its spatial coordinates, its likely behavior, and so on. For this reason, it has been convincingly argued that the information about the object is processed as relevant to the subject's well-being. When it comes to the secondary or social emotions -- Damasio speaks for example of hearing the news that a friend is dead -- the situation is even worse, for one wonders how it is possible to feel grief, pride, or shame at all without having in the mind/brain the evaluative representations not only of objects, but of the complex situations these emotions are about.

On second thought, however, philosophers and psychologists fond of the "aboutness" of the emotions should be reminded that no one seems to know how a neural representational pattern might be built into an emotional state, or vice versa (see, e.g., LeDoux, 1996). If Damasio and other neuroscientists are right, it may well be that in the brain the relation between representations and emotional states is not "conceptually internal", but rather externally associative. Some philosophers of mind and cognitive psychologists (though not memory theorists) almost always shun the old-fashioned relation of association, but perhaps neuroscience can teach us that all there exists in nature is associative connections through neural pathways. This is food for thought for those of us who believe that the mind ought to be thoroughly naturalized.

My second observation is concerned with Damasio's concept of feeling. This is described as a "sensing" of the consequences of activation in some brain regions, but only "provided the resulting collection of neural patterns becomes images 'in mind'" (p. 79). This is a sudden and surprising switch to the realm of the mental (there has been some warning, to be sure; on p. 42 we are told that "feeling is the private, mental experience of an emotion.") The switch is not so problematic as it stands: feeling might be physical and psychological at the same time, or perhaps it is just one and the same thing. What is problematic is the "private", because privacy presupposes an owner, a subject, or more accurately a higher-order state of the mind-brain which has exclusive "access" to, or can be uniquely affected by, a lower-order state. If feelings are private, then there must be a higher-order level of consciousness which can monitor them -- this is precisely what Damasio will argue. But surely there are unconscious feelings which are not monitored by a second-order consciousness, hence contrary to Damasio not all feelings are necessarily accessible even to the person whose feelings they are.

3. Is Consciousness Necessary for Emotion?

Damasio believes that consciousness comes in degrees, and one of the distinctive features of his account is the proposal to withhold the title of "conscious" altogether from a number of states which are nevertheless "mental" and are often ascribed to consciousness of the first order. To make this clear, he begins where not even mentality is actually present; there are states, for example, in which even the most elementary wakefulness is lacking. These include coma, deep anesthesia, and normal dreamless, non-REM sleep. Patients don't respond to stimuli, lack muscular tone, and are incapable of the slightest attention to events in their surroundings, let alone of self-consciousness. Even background emotions are missing.

Mentality begins with states of wakefulness in which attention and purposeful behavior are still fleeting and unfocused. These include the "persistent vegetative state", similar to coma but with cycles of sleep and wake, "absent seizures" short of convulsions in epileptic patients, Alzheimer's disease, hypnosis, and... the proverbial absent-mindedness of college professors. In these cases there are no (behaviorally detectable) emotions, either.

Finally, there is an interesting class of borderline states characterized by wakefulness and a minimal degree of attention and purposeful behavior. These are the epileptic automatisms following an absent seizure, during which a patient may even walk out of the door and in the street, but is clearly not self-conscious; and the "akinetic mutism" of patients who have suffered a stroke, may notice their surroundings, but stare in the void for months and remember nothing in the aftermath. In neither case are there any signs of emotion. Now to neuropathology. Here is a mapping of the key symptomatologies to the relevant brain lesions, in order of decreasing severity:

- i) No wakefulness, no emotion: Upper brain stem, hypothalamus, thalamus.
- ii) Wakefulness with unfocused attention and non-purposeful behavior, no emotion: Upper brain stem, hypothalamus, thalamus.
- iii) Absent seizures, no emotion: Thalamus, or anterior cingulate cortex.
- iv) Wakefulness with minimal attention and some purposeful behavior: Cingulate cortex, medial peri-cingulate cortex, basal forebrain, thalamus.

The first two classes of patients suffer from lesions in brain stem, hypothalamus, and thalamus, of differing gravity; not surprisingly, emotion is lacking in all of these states (see paragraph (1) above). Absent seizures are caused by lesions in the thalamus and the anterior cingulate cortex, yet emotion is still not there. The same holds for akinetic mutism and epileptic automatisms, for which the dysfunction is in the cingulate cortex, the basal forebrain, the thalamus, and the medial peri-cingulate parietal cortex. These are structures that we have found to be responsible

for the sensing, or feeling, of the consequences of emotional activation, and it is probably for this reason that Damasio suggests that "absence of emotion is a reliable correlate of defective core consciousness" (p. 100) -- his term for momentary introspective self-consciousness. In other words, the basic subcortical structures responsible for emotion seem to be necessary but not sufficient for it: consciousness would also be necessary. Indeed, patients who do have core consciousness -- amnesics, for example -- have both background and primary emotions, whereas a more sophisticated extended consciousness seems to be required for the secondary kind.

The coupling of emotion and consciousness, however, remains puzzling. It defies the elementary intuition that a subject may, say, react emotionally to a movie, or even scream with rage, without being aware *that* he or she is doing so. Damasio seems to have doubts, too. He acknowledges that emotions can be triggered unconsciously by unattended thoughts or imperceptible aspects of body states -- on p. 42 he explicitly claims that "the basic mechanisms underlying emotion do not require consciousness, even if they eventually use it" -- but then insists on the requirement of core consciousness. (In *Descartes' Error*, too, he admitted that skin conductance responses characteristic of emotion may fail to be perceived, and also that the "body loop" may perhaps be replaced by an "as if" body loop, completely within the brain; pp. 156, 209). The reason for the insistence on core consciousness would be that "both emotions and core consciousness require, in part, the same neural substrates, and that strategically placed dysfunction compromises both kinds of processing" (p. 100). This is true -- in part -- but are we not confusing the basic mechanisms with the sensing of their activation? The cases of absent seizures, epileptic automatism, and akinetic mutism remain puzzling, but then we need to be told how to accommodate ordinary intuitions (which may well be discarded if necessary). Besides, if the clinical evidence shows that core consciousness is necessary for emotion, and we follow Damasio in his thinking that emotion is necessary for consciousness, why would we need a distinction between the two? Isn't it perhaps the case that there can be emotion without consciousness, and consciousness (of something else) without emotion? It is unfortunate that a strong neuropathological approach should not yield the conceptual clarity that one would expect -- which by no means implies that such an approach should not be pursued. The problem calls for further exploration.

4. The "Proto-self" and the "Core Self."

According to Damasio, the key to self-consciousness does not lie in the cognitive processes traditionally studied by neuroscientists. The key, rather, lies in the "proto-self", a novel concept referring to "a coherent collection of neural patterns which constantly map, moment by moment, the state of the physical structure of the organism in its many dimensions" (p. 154). When sensory devices perceive external objects, the organism must constantly adapt to perceptions by adjusting the position

of the lens and the pupil, and the muscles of the head, the neck, and the trunk; these adjustments are detected by the somatosensory areas which constitute part of the proto-self. Most importantly, the proto-self receives neural and hormonal signals from visceral changes, such as emotional reactions. (Perhaps we have found the "location" of feeling.) It is one of Damasio's central theses that there is no such thing as "cold", disinterested perception; every perception is relevant to the well-being of the organism, and is detected as such by the proto-self. The proto-self is not conscious -- first-order consciousness does not have a place in the theory; yet for Damasio it constitutes the biological precedent of the self. The next step will be to see why this is so, but before that we should examine the neurology of what we are discussing.

The relevant structures are the brain stem, the hypothalamus, and the somatosensory cortices (insular cortex, S2 cortices, medial parietal cortices), especially in the right hemisphere where spatial perceptions are mostly located. Notice that structures responsible for memory (the hippocampus) or abstract thought (pre-frontal cortices) are not required -- this however would exclude emotions caused by memories. Why the proto-self should include the brain stem and the hypothalamus, which it seems to monitor, is not clear. The answer is probably that these structures are both activated in emotion, and contribute to monitor somatic events induced by their activation. This impression is confirmed by a detailed and fascinating discussion, later in the book, of recent research on the reticular formation within the brain stem; there Damasio actually seems to identify the proto-self with some acetylcholine and monoamine nuclei known to have homeostatic, emotional, and also wake-sleep functions. But then again, consciousness too appears in this discussion, whereas one would have thought that we were not at that level yet.

What, then, is core consciousness? For Damasio, it is a "non-verbal account" of how the organism's state is affected by the processing of an object. The exposition, to be sure, is not always lucid, but the main ideas are clear enough. The brain forms neural maps (a) of an object, whether perceived or remembered -- this is where object-representations come back into the picture -- and (b) of the organism -- this collection of maps is the proto-self. Both (a) and (b) are called first-order maps. Now, the sensory-motor maps pertaining to the object -- (a) -- cause changes in the maps pertaining to the organism -- (b) -- because the maps pertaining to the organism register the organism's responses to the object, or more accurately to the map of it. Finally, the changes in the organism-maps, and the object-maps as well, are in turn represented in, or by, second-order maps (c), which constitute the core self or core consciousness. All maps are neural patterns, and all can become "mental images", though we are not told how. So far, so good, at least if the reader manages not to get confused by expressions such as "The mapping of the object-related consequences occurs in first-order maps representing proto-self and object" (p. 170) -- wasn't the proto-self the map which represents changes in the organism? In several passages Damasio uses "represent" and "representation" in the sense of "display", or even "constitute", rather than in the familiar sense of a relation

between a state or object A and another object B. This in principle would be quite all right (those who know what representing means please raise their hand), but then the novel usage should be consistent (cf. "the organism in the hypothesis is represented by the proto-self," p. 170, where the usage is standard).

Where is the second-order neural pattern located? Damasio warns us that it is a mistake to think of a single consciousness center -- he is neither a phrenologist nor a "Cartesian materialist" (Dennett, 1991; Dennett and Kinsbourne, 1992) -- and suggests instead that there is a parallel simultaneous activation of different structures. Not many, however. The most important one is the familiar cingulate cortex (especially its anterior sector). This is not a novelty, to be sure. Damasio already pointed to the cingulate cortex in his 1994 book, and in any event the idea has been around for a long time: J.W. Papez postulated its function of emotional consciousness as early as 1937 (Papez, 1937). The good news is that we are now offered a rather detailed discussion of this puzzling structure which seems to have myriad functions, both sensorial and motor-related, and Damasio also ventures some hunches about the function of its posterior sector, which is not well known (p. 263). Other structures responsible for core consciousness are the thalamus and the superior colliculi (a tiny structure near the brain stem, sometimes regarded as the seat of consciousness). Damasio notably excludes the pre-frontal cortices, granting them at most a role in working memory and extended consciousness. This is indicative of his overall conception and aim, which is to emphasize the function of a transient and ephemeral core consciousness, independent from language, memory, and reason, and yet capable of what is sometimes called reflexivity. He aptly reminds us that in the history of Western thought higher levels of consciousness, linguistic performance, and rationality have been explored before and more thoroughly than the relatively simpler states which gave rise to them.

The story, then, is roughly the following. Sensory cortices (or the hippocampus in the case of memory) map an object, and normally also activate emotion-related structures, at least the brain stem, the hypothalamus, and the amygdala. These structures do three things: first, they trigger autonomic reactions in the soma; second, they send neural messages to other brain areas; and third, together with the somatosensory cortices, they also map or "represent" the somatic reactions that they have induced (together with the somatosensory areas they constitute the proto-self). Finally, the anterior cingulate cortex, the thalamus, and possibly the superior colliculi map both object-maps and the ever-changing organism-maps, and this is core consciousness. Put in this way, this account suggests a relatively neat distinction between orders of mapping or representation: the basic structures seem to be necessary and sufficient for emotion, but are not sufficient for the consciousness of it, which requires the activation of other areas. But this is not exactly what Damasio says; for if the cingulate and other cortices are already involved at the beginning of the process -- the cingulate seems to be a "massively somatosensory structure" in its own right (p. 261) -- everything happens (almost) everywhere and there is no clear second-order mapping to speak of.

Other observations can be suggested. At first sight, Damasio's view seems to be a variant of what is often called a representational theory of consciousness. According to this theory all states of consciousness, including those traditionally classified as "phenomenal", like sensations, are representational (see, e.g., Dretske, 1995). At least in the philosophical literature, the issue then arises whether or not first-order conscious representations of external objects can account for, or explain away, the so-called qualitative properties of conscious experience. Damasio does engage in a brief discussion of Frank Jackson's famous story of Mary the neuroscientist, who knows everything about color processing but has never experienced color (Jackson, 1982), and vigorously rejects his claim that knowledge without experience shows the irreducible incompleteness of knowledge (surprisingly, and incorrectly, he also puts the arch-rivals Searle, 1992, 1998, and Dennett, 1991, in the same camp on this issue). But this is not really his main preoccupation. He discusses neither the qualia of object-maps -- if any exist -- nor those of the organism-maps which are at the center of his account. In other words, he may be regarded as a representationalist by default. The fact is that Damasio focuses throughout his work on second-order consciousness, which brings him to say for example that "the experience of a particular stimulus, including color, depends not just on the formation of an image [a mental pattern], but also on the sense of self in the act of knowing" (p. 308). One wonders, however, if the problem does not reappear at the second-order level: Is a representation of an object-map the same thing as a representation of an organism-map? In other words, are perceptions the same thing as sensations? Damasio is silent on this.

Moreover, if it is all a matter of registering physical changes -- even at the level of second-order core consciousness -- why talk of *representations*? In his Glossary, Damasio offers what he calls a "conventional and transparent" definition of the term as "a pattern", whether neural or mental, "that is consistently related to something" (p. 320). Unfortunately, this is much too simple. A naturalistic account of this crucial semantic term, which Damasio also abundantly employed in *Descartes' Error*, is more than welcome, but there are well-known difficulties with what philosophers call a co-variational account. Briefly, a representation cannot simply be an indicator or effect of its cause, because it may be *mistaken* (think of phantom-pains, blindsight, various agnosias, or indeed of any illusion or hallucination, which incidentally often trigger emotions about misrepresented or even non-existent objects). From this it follows, at least according to one possible account, that we need to define the content of a representation in terms of what *would* be its cause in normal circumstances, and normal circumstances are those in which a correct representation has a function, defined by natural selection (Millikan, 1984; Dennett, 1996). How this account can be extended to representations of artifacts and cultural objects in general (e.g., institutions) is another, mostly unexplored, story, but Damasio does not even begin to discuss the central thesis. This is disappointing, given his biological orientation. The unfortunate result is that second-order consciousness is surprisingly described as a device which regularly and correctly represents first-order maps which in turn cannot be mistaken.

5. How Many Levels of Consciousness?

In my discussion so far I have followed Damasio in his thinking that there are two "orders", of which the proto-self and representations of external objects are the first, and core consciousness is the second. This is more or less clear, despite some statements that perhaps are slips of the pen. For example, at one point we are told that "we are *not* conscious of the proto-self" (p. 174). This presumably means that the proto-self can sometimes monitor the state of the organism without being monitored in turn. But of course, if core second-order consciousness is to be possible at all, we'd better at least be *able* to be conscious of the proto-self, since second-order consciousness is *about* first order states. So there are two levels or orders so far: the proto-self, which is the cause, and the core self, which is the effect.

At this point, however, Damasio claims that "we are conscious of the core self" (p. 174). The reader begins to suspect that he has fallen into the familiar trap of the double sense of "conscious", one of which is "active" (conscious of), while the other is "passive" (conscious as monitored by another state). Perhaps what he wants to say is simply that the *core self* is (actively) conscious. If not, then who is "we"? Isn't Damasio adding yet another, third level? The trouble here is that the use of first person pronouns often conceals a level of reference (or causation, for that matter). Third person discourse makes this perspicuous: Bob is conscious of his core self which is conscious of his proto-self, which represents the organism. Better still: There is an area or pattern in Bob's brain which is affected by his core-self... etc. Now there is nothing wrong with nested intentionality (or causation), but then Damasio owes us an explanation of how the third level, or third-order map, works, and where it is located. One possible answer could lie in the "autobiographical self" or extended consciousness, which is said to be constituted of memories, and hence presumably also of memories *of* core self states (extended consciousness, we are told, requires the operation of temporal and frontal higher-order cortices, as well as of the amygdala; see p. 220). On the same page 174, however, Damasio speaks of an autobiographical self "of which we are conscious." This is again plausible, but it introduces a *fourth* level, of which Damasio does not seem to be aware. Indeed, one major weakness in his whole theory is that the multiplication of levels gets somehow out of hand. This is apparent when he returns to feelings, and to feelings of feelings, towards the end of the book.

6. A House of Cards?

We already know what feeling is. It is the mental experience of an emotion, and Damasio is quite clear about this point: this is not an "emotional experience", but a separate experience which happens to be of, or caused by, an emotion (remember

that an emotion, on his account, is a pattern of chemical and neural responses). So a subject feels that she or he has an emotion. Accordingly, we read that "feeling an emotion consists of having mental images arising from neural patterns which represent the changes in body and brain that make up an emotion" and that this requires second-order representations necessary for core consciousness (p. 280). This is clear enough. The trouble is only that at this point, within four pages (281-284), the whole construction explodes like fireworks in a cascade of concepts and relations that are an impossible challenge for even the most attentive and sympathetic reader.

The storm appears over the horizon where Damasio begins to speak of "knowing that we have that feeling, 'feeling' that feeling" (p. 280), which for all intents and purposes should denote a third-order state distinct from feeling. Later on the same page, he declares that "the proto-self, feelings of emotion, and the feeling of knowing feelings emerged at different points in evolution and to this day emerge at different stages of individual development" (*ibid.*). The principle of charitable interpretation suggests that the cryptic "feeling of knowing feelings" should be read as "feeling-of-knowing feelings", where the "of" is not representational or causal: higher-order feeling just is knowledge of lower-order feeling. Anyhow, the impression is still sustained that we are talking about three levels not two. The same impression is confirmed on p. 282: "When those images [constituting feelings] are accompanied, one instant later, by a sense of self in the act of knowing... they become conscious. They are, in a true sense, feelings of feelings." Well, you figure it out. One more time: "I am suggesting that 'having a feeling' is not the same as 'knowing a feeling,' that the reflection on feeling is yet another step up" (p. 284).

I conclude that on Damasio's own account a third-order level seems inescapable, but unfortunately he nowhere acknowledges this. The worst consequence is that after learning from him how to think of consciousness in terms of pretty well defined brain regions, we are left in the dark about the localization of the third order as distinct from the second and the first. Notice, by the way, that a third-order level is very plausible from a psychological point of view: one can approve of one's shame of one's fear (ethics lurks in the wings). Hence either we add a third order above core consciousness -- it might be extended consciousness -- or we take core consciousness to be the third order and then find a second order corresponding to the feeling of emotion (perhaps the proto-self is a candidate).

The situation is clearly not satisfactory, and it is worth wondering why. Damasio, like other neuroscientists, typically begins with a strictly neurobiological explication of the psychological concept of emotion, and does not attribute to these neural processes any representational or phenomenal property; then he proceeds to explicate the psychology of feeling and consciousness by investigating neural projections to structures of the second and perhaps third order, which are said to represent both external objects and first-order reactions. Neuropsychology begins, indeed, with psychological categories, although it must leave open the possibility that neurological findings might force us to revise or abandon psychological

hypotheses, like for example the representational nature of emotions. The result is often a fruitful interaction between levels of analysis, but *only if* there is a minimum of consistency and transparency in the psychological picture that is "tested", and this we have found wanting. Consequently, it is difficult to tell whether the psychological story is confirmed by the neurological evidence, or whether it simply clashes with it and must be discarded. This leaves the reader with a feeling of dissatisfaction -- whatever that means.

7. Conclusion.

Antonio Damasio has offered a thought-provoking view of consciousness centered on feelings, and feelings of feelings, of neural and chemical reactions that he calls emotions. We have followed him through his argument, and we have found that there is much to learn from neuropathology about different levels of consciousness, while there is more work to do on psychological levels and the corresponding neural structures. The time has come, however, to wonder about the limits of the whole project.

There is no doubt that the overwhelming majority of research in cognitive neuroscience and psychology has been oblivious to emotional processes, as if the brain were an "epistemic engine" and nothing else. But it does not take a seventeenth-century rationalist to see that many first- and second-order representations, at least in some contexts, are merely epistemic and emotionally irrelevant -- think of the copy-machine in your office. From this it follows that second (or third?) order structures such as for example the cingulate cortex, which seems to be crucial to the consciousness of emotion, would have to be shown to be equally crucial to non-emotional core consciousness as well; or else, we still wish to know what makes sheer epistemic consciousness possible. Only then will we have a viable theory of consciousness as a whole.

References

- Damasio, A. (1994). *Descartes' error: emotions, reason, and the human brain*. New York: Avon Books.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little Brown.
- Dennett, D. (1996). *Kinds of minds*. New York: Basic Books.
- Dennett, D., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, 15, 183-247.

Descartes, R. (1649/1984). *The passions of the soul*. In J. Cottingham, R. Stoothoff, D. Murdoch (Eds.) *The philosophical works of Descartes*. Cambridge: Cambridge University Press.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge: MIT Press.

Ekman, P. (1992). Facial expressions of emotions: New findings, new questions. *Psychological Science*, 3, 34-38.

Frijda, N. (1986). *The emotions*. Cambridge: Cambridge University Press.

Jackson, F. (1982) Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127-136.

James, W. (1884). What is an emotion? *Mind*, 9, 188-205.

Lazarus, R. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

LeDoux, J. (1996). *The emotional brain*. New York: Simon and Schuster.

Levine, (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64, 364-361.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge: MIT Press.

Papez, J.W. (1937). A proposed mechanism of emotion, *Archives of Neurology and Psychiatry*, 79, 217-224.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge: MIT Press.

Searle, J. (1998). *Mind, language, and society*. New York: Basic Books.