

On the Human

a project of the National Humanities Center

Doing, Feeling, Meaning And Explaining

by

Stevan Harnad

Chaire de recherche du Canada

Institut des sciences cognitives (ISC)

Université du Québec à Montréal

Montréal, Québec, Canada H3C 3P8

&

School of Electronics and Computer Science

University of Southampton

Highfield, Southampton

SO17 1BJ UNITED KINGDOM

ABSTRACT: It is “easy” to explain doing, “hard” to explain feeling. Turing has set the agenda for the easy explanation (though it will be a long time coming). I will try to explain why and how explaining feeling will not only be hard, but impossible. Explaining meaning will prove almost as hard because meaning is a hybrid of know-how and what it feels like to know how.

We can reduce just about everything that cognitive science needs to explain to three pertinent anglo-saxon gerunds — doing, feeling, and meaning — plus a fourth, which is explaining itself.

There’s *doing*, which covers everything that people and animals are able to do (all of their “know-how”). That’s not just moving around; it includes recognizing and manipulating objects in the world and exchanging strings of words (talking, language) about them ([Harnad 2010](#)).

Then there’s *feeling*: it *feels like something* to do most of the things people and animals are able to do (while they’re awake). It feels like something to see, recognize and manipulate an object.

Then there’s *meaning*: The strings of words that people (not animals) exchange *mean* something.

And, last, there is the problem of *explaining* all of the above: explaining how people can do what they can do; explaining how people can feel; and explaining how words can mean.

It has become fashionable lately to call the problem of explaining *doing* the “easy problem” — compared to explaining feeling (i.e., consciousness), which is the “[hard problem](#)”.

The easy problem is easy in the sense that although cognitive science cannot yet explain how people and animals are able to do what they can do, there’s no reason to expect that eventually it will not succeed. Alan Turing set out the method in 1950: Design a robot that can do anything and everything that a normal human being can do, and whatever turns out to be the successful causal mechanism inside that robot — the one that gives it the capacity to do whatever we can do — will be an explanation of how and why we ourselves are able to do what we can do ([Harnad 2008](#)).

People usually reply: But I’m not a robot, and I’m not interested in how robots can do things. I want an explanation of how *I* do things. (That usually means: I want an explanation of how my brain does things.) Fair enough. For those who are unsatisfied with a causal mechanism that can “merely” do anything and everything a normal human being can do — even a “Turing robot” that can pass the Turing Test, passing as one of us, indistinguishable from a real human being for an entire lifetime based only on what it does and says — the causal mechanism can be further elaborated so it also “does” everything the brain does, internally as well as externally ([Harnad 2011](#)). Externally, of course, the brain does what our bodies do. But internally there are neurons and connections and patterns of activation, even chemical “doings.” A “Turing biorobot” must do both.

Now once we have a Turing biorobot — its doings, external and internal, indistinguishable from our own — the “easy

problem of explaining doing is solved. But what about feeling? Would the Turing robot or the Turing biorobot feel? Unfortunately, unlike doing, feeling is not something that can be observed by anyone other than the feeler. And the Turing robot, if it is indeed indistinguishable from us, would of course behave exactly as if it feels. And if asked, it would reply exactly as any of us would reply: "Of course I feel! What a question!" This is called the "other minds problem": The only one I can be sure feels is myself. With others, I have to infer it from what they do and say (including how they move and how they look and sound: facial expressions, tone of voice).

The other-minds problem, however, is not the "hard problem," although it is related to it. Even though we can't know for sure whether other people feel, it's just about as probable that they do feel as the fact that apples always fall down, not up (we can't be sure about that either, but it's close enough): People are pretty good at "mind-reading." So that's why we don't worry about whether other people really feel, or just act and talk as if they do.

Now the Turing robot is indistinguishable from the rest of us on any of these external cues; the Turing biorobot is not even distinguishable on internal cues (although we don't normally invade people's brains to determine whether they feel!). But we do know that Turing robots are synthetically made rather than natural, so the uncertainty about whether or not they feel is greater than it is with our fellow human beings (greater even than with our fellow animals, except perhaps the ones that are the most unlike us) ([Harnad 2001](#)).

But the "hard problem" is not that greater uncertainty about whether a Turing robot or biorobot feels. Let's suppose it does feel. That doesn't help, because that still doesn't explain *how and why* it is able to feel, if it does feel. In the case of explaining its know-how — the causal mechanism that successfully generates its ability to do what it can do, indistinguishably from the rest of us — the explanation is complete, and accounts for anything and everything the robot can do, whether or not it can feel. If it does feel, that's nice; but, unlike the doing, *the feeling is not explained by the causal mechanism of the doing*.

We might pause and consider just how hard a problem this is. We feel. For example, we feel pain when our tissues are injured. The temptation is to say that we need to feel the pain otherwise we would not notice the tissue injury and we would not do what needs to be done about it. But doing is doing. If something needs to be done, why is it not enough to have a mechanism that, when it detects tissue injury, sees to it that what needs to be done is done — withdraw the hand from the fire, avoid fire in future, remember, learn, compute and even *say* whatever needs to be done — without bothering to feel anything at all? What's the bonus from feeling something? What causal role does feeling fulfill, that doing alone does not? Our Turing robot can do everything that needs to be done, whether it feels or not. If it does feel, it remains to explain what causal role the feeling itself is playing. And that's the hard problem ([Harnad 1995](#)).

There would be an easy solution if there were psychokinetic (mind-over-matter) forces in the world, alongside ordinary electromagnetism, gravitation and atomic forces. Then feelings could be an independent further force, and their causal role would be measurable and explainable. But there aren't any psychokinetic forces (despite generations of parapsychology experimentation seeking to detect them). And even aside from the fact that there's no evidence for psychokinesis, not only are the known physical forces already enough to get anything that needs doing done as well as explained (whether the doings are those of a galaxy, an atom, a steam engine, an organ or an organism), but there isn't even any causal room left for any forces beyond the known ones.

So whereas explaining doing is easy, explaining feeling is hard (perhaps even impossible). Now let's move on to meaning: Let's consider written words (though we could just as well have considered spoken words, or words in a gestural language): Written words have a graphical "shape": They look like something. And, in addition, they also mean something. Now a Turing robot, like us, could detect the shape of a word: could read it, speak it, point to its referent (if it has a concrete referent, like an apple) or describe its referent (if it has a more abstract referent, such as "truth") ([Harnad 1990](#)).

But all of that is just doing. Is meaning, too, just doing — as in the ability to point to what a word refers to (an apple) as well as to describe the word's sense (a round, red fruit)? I suggest that it is no more true that meaning is just doing than that seeing an apple just amounts to the ability to identify an apple. It *feels like something* to see an apple when you are looking at one. And it also feels like something to *mean* an apple when you are talking (or thinking) about one.

The fact that differences in meaning are also felt differences is even clearer with words that have double meanings: the word "pound" can be used in the sense of a unit of weight or in the (British) sense of a unit of currency. The words sound the same, but they have different meanings. Context presumably sorts out for the hearer what the speaker means when he

says he's "lost a pound."

But what about when I'm the one saying it, and all I say is "I've lost a pound"? Not only does it feel different (to me) to say and mean "I've lost a pound" in the sense of losing money, compared to what it feels like to say it in the sense of losing weight, but such differences in feeling — subtle though they are, and hard to describe — are not just differences in what I do, or can do, or am disposed to do: They are also differences in what I feel: *It feels like something to mean something*. It also feels like something to think, believe, understand or doubt something.

Most philosophers reject this conclusion. They suggest that meaning, thinking, believing, understanding or doubting something is not the same sort of thing as seeing, hearing, touching or tasting something, nor even like feeling a migraine or a mood. So don't ask a philosopher; instead, ask a psychophysicist. Psychophysicists are the ones who specialize in measuring sensations and the detection of differences: Does this look (sound, smell) brighter (louder, stronger) than this? They have even narrowed it down to the "just-noticeable-difference," or JND, which is the smallest difference between two inputs that a person can tell apart.

Psychophysicists don't philosophize; they just measure what differences people can and cannot tell apart — in other words, what people can *do*. But if you ask psychophysicists how people make same/different judgments, they will of course tell you that it's based on whether things feel (look, sound, taste, smell) the same or different. You can't do much psychophysics on someone who is fast asleep, not feeling a thing.

Well, in psychophysical terms, we are making same/different judgments on the basis of differences in meaning (not sound!) constantly, in our discourse. Insofar as measurement is concerned, JNDs in semantic space would look pretty much the way JNDs do in sensory space. It would be surprising, then, if sensory differences were felt, whereas semantic differences were unfelt. That would make talking and thinking much more like "blind-sight" — in which patients with certain kinds of brain damage report that they can no longer see at all, yet they are somehow still able to distinguish things presented to their (intact) eyes (Overgaard 2011). The current consensus is that blind-sight patients are still feeling *something* — even if it is only the movement of their eyes, which is controlled by an eye-movement control system that is also still intact in their brains — and they are using that feeling of involuntary movement (or the felt urge to move) as the basis for telling (some) things apart.

But "blind-semantics" would be more like "speaking in tongues," with words coming into our ears and going out of our mouths as if they were a foreign language that we did not understand (rather the way Searle 1980 describes it, in the Chinese Room). Yet we know that we can sense what we are meaning as surely as we can sense what we are seeing. And "sensing" is just a synonym of "feeling," here.

Why do (most) philosophers think that meaning — unlike sensing and emoting — is unfelt? It has to do with the "hard" problem, again: It's so hard (perhaps impossible) to explain how and why we feel that it seems like a good idea to try to wrest as much as possible of cognition from the clutches of feeling, to get it over onto the "easy" side of the ledger — doing — the side that we have some hope of explaining. And, as we've already noted, talking to one another is certainly something we do. But distinguishing and manipulating things through viewing and touching is also something we do. And we all know that different things look different (up to a JND). Surely the differences among the things we say and mean are not just differences in the shapes of the words (or what they sound like)? Rather, differences in meaning are felt differences too (O'Callaghan 2011, Strawson 2011).

So the "hard" problem of explaining how and why we feel is even more pervasive than sensory and emotional experience. Semantic sense is afflicted with it too.

I would like to close by suggesting that the hard problem is not a metaphysical one, at least not for cognitive science. It's surely true that the brain causes both doing and feeling (and hence also meaning), *somehow*. The problem is explaining how — and, even more problematic, *why*? With doing, it's easy to explain how and why we can do what we can do. With feeling it's hard, if for no other reason than that *doing alone already does the job*: it's enough to explain what kind of Darwinian survival engines organisms are, i.e., by what causal process we evolve or learn the ability to do what needs to be done for our survival, reproduction and lifetime success (such as it is). The Turing robot (or biorobot) will be fully explained by the explanation of the causal mechanism that generates its ability to do what it can do. Even if the robot does feel, that causal explanation will not explain how it can feel, let alone why. The causal mechanism will be equally compatible with the presence or absence of feeling.

The challenge to my commentators, then, is to use whatever actual facts you know about our know-how and the internal

mechanisms that generate it — or whatever speculative evidence and mechanisms you think could turn up in the future — in such a way as to sketch how it could ever be explained *how and why we feel rather than just do*. If (as you should) you include our language capacity and use in our know-how (Harnad 2007), then please also address the problem of meaning: Why does it feel like something to understand someone's words, rather than simply causing unfelt internal processes, that in turn cause further words and other doings on our part?

"Causing" is evidently our fifth pertinent gerund, and feeling seems to be causal explanation's nemesis.

References

- Harnad, Stevan (1990) The Symbol Grounding Problem *Physica D* 42: 335-346.
- _____ (1995) Why and How We Are Not Zombies. *Journal of Consciousness Studies* 1:164-167.
- _____ (2001) Spielberg's AI: Another Cuddly No-Brainer.
- _____ (2007) From Knowing How To Knowing That: Acquiring Categories By Word of Mouth. Presented at Kazimierz Naturalized Epistemology Workshop (KNEW), Kazimierz, Poland, 2 September 2007.
- _____ (2008) The Annotation Game: On Turing (1950) on Computing, Machinery and Intelligence. In: Epstein, Robert & Peters, Grace (Eds.) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer
- _____ (2010) From Sensorimotor Categories and Pantomime to Grounded Symbols and Propositions. In: *Handbook of Language Evolution*, Oxford University Press.
- _____ (2011) Minds, Brains and Turing. *Consciousness Online*.
- O'Callaghan, C. (2011) Against Hearing Meanings. *The Philosophical Quarterly*. DOI: 10.1111/j.1467-9213.2011.704.x
- Overgaard, Morton (2011) Visual experience and blindsight: a methodological review. *Experimental Brain Research* 209(4): 473-9
- Searle, John R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (3): 417-57
- Strawson, Galen (2011) *Cognitive phenomenology: real life*. In T. Bayne & M. Montague (eds.): *Cognitive Phenomenology*. Oxford University Press
- Turing, A.M. (1950) Computing Machinery and Intelligence. *Mind* 49: 433-60

32 comments to Doing, Feeling, Meaning and Explaining

Judith Economos

April 19th, 2011 at 8:40 am

I insist, incorrigibly, that it does not feel like anything, and does not feel, when I wonder idly if it will rain next Thursday, and perforce it does not feel different from wondering if it will rain next Wednesday. I know that I have wondered and know which day I wondered about, but not, I solemnly swear, by any feelings. I just know, and this seems a plentiful and elegant sufficiency.

This is not abstract speculation; this is not philosophy; this is testimony. I am in possession of my faculties and have been known to understand and write English pretty well. Does this not count? Is it somehow inconclusive, that we are still arguing about feeling our thoughts?

.....

Good evening, Stevan,

That I know what I am thinking is setting me apart from a computer or a zombie that does not know what it is thinking and therefore is only "thinking". You see, I do understand that much. It just seems to me important not to confuse this with feeling. but all the words I might use, like "aware" and "conscious", are bad words, dishonest and

weaselly. It is true that those words are vague, but so is Feeling, as you use it, and the problem is that what goes on in our minds is private to us and not really shareable. We have to communicate it to each other with words whose meanings to the other can at best be guessed at by analogy with states we think are comparable, working from the most obvious behavioristic criteria to the more delicate shades of inward state, events, processes, efforts, and, er, well, awarenesses. Since we cannot directly compare your mental target with mine, we just can't know if we are really on about exactly the same thing, or even if there is such a thing in my mind as there is in yours. That is why I am comfortable in my intransigence about what I feel or don't feel, and why I think argument is futile. Each of us is (as it were) sitting in a closed room considering objects for which we have never had a common name, nor common adjectives, and trying to see if the other has similar objects in his room.

Stevan Harnad

April 20th, 2011 at 5:02 pm

INCORRIGIBILITY AND IMAGELESS THOUGHT (Reply to Judith Economos)

I think the debate about whether there are unfelt thoughts may be a throwback to the introspectionists' debates about whether there is "*imageless thought*."

The big difference is that the introspectionists thought that an empirical methodology was at stake, one that would allow them to explain the mind. We now know that there was no such empirical methodology, and no explanation of the mind was forthcoming, not just because introspection is not objective, hence not testable and verifiable, but because it does not reveal how the mind works.

(This conclusion was partly reached because it was not possible to resolve, objectively, the rival claims of introspective "experimentalists" as to whether imageless thought did or did not exist. The outcome was an abandonment of introspectionism in favor of behaviorism, until it turned out that that could not explain how the mind works either, which led to neuroscience and cognitive science.)

The difference here, is that we are debating about whether or not there are unfelt thoughts while at the same time *agreeing* that introspection is neither objective nor explanatory.

The disagreement is, specifically, about what it means to say that we are conscious of something, when that being-conscious-of something does not feel like anything.

The reason I keep calling "consciousness" and "being conscious of" something *weasel-words* is that if things are going on in my mind that somehow *are* different yet do not *feel* different then it is not at all clear how I am telling them apart, or even how or what I am privy to when I say I am "conscious of" them despite the fact that I don't feel them in any way.

It seems to me that the notion of "unconscious thought" and "unconscious mind" is as arbitrary and uninformative as "unfelt thought" — or, put in a less weaselly way that bares the incoherence: unfelt feelings. Among other things, this makes it more obvious that "thought" and "thinking" are themselves weasel-words (or, at best, phenomenological place-holders). For apart from the fact that they are going on in my head, it is not really clear what "thoughts" are: we are waiting for cognitive science to tell us! (We keep forgetting that "cognition" just means thinking, knowing, mentation — weasel words, all, or at least vague if not vacuous until we have a functional — i.e., causal — explanation of what generates and constitutes both them and, more importantly, what we can do with them, apart from thinking itself.)

Now explaining how we can do what we can do is these days called the "easy" problem. What, then, is the "hard problem," and what is it that makes it hard?

My own modest contribution is just to suggest that (1) the hard problem is to explain how and why thought (or anything at all) is felt, and that (2) it is hard because there is no causal room for feeling in a complete explanation of doing.

So where is the disagreement here? Let us assume we can agree that explaining *unconscious* thought — i.e., those things going on inside the brain that play a causal role in generating our capacity to do what we can do, but that we are not conscious of — are *not* hard problems but easy problems. They may as well be going on inside a toaster. The rest is just about how they generate the toaster's capacity to do whatever it can do. And of course we are not privileged authorities on whatever unconscious "thoughts" may be going on in our heads, because we are not conscious of them.

So all I ask is: What on earth do we really mean when we speak of having a thought that we are conscious of having, but that it does not feel like anything to have!

You (rightly) invoke "incorrigibility" if I venture to doubt your introspections, but, to me, that justified

insistence that only you are in a position to judge what's going on in your mind [as opposed to your head] derives from the fact that only the feeler can feel (hence know) what he is feeling: what the feeling feels like. But apart from that, it seems to me, there are no further 1st-person privileges. One can't say: I don't feel a thing, yet I know it. On what is that privileged testimony based, if it is not the usual eye-witness report? "I didn't see the crime, but I know it was committed?" How do you "know"? In the case of thought, what does that "knowing" consist, if not in the fact that you feel it (and it feels true)?

There is no point referring to objective evidence here. I can "know" it's raining in the sense that I say it's raining and it really is raining. I have then made a true statement, just as a robot or a meteorological instrument could do, but that has nothing to do with the mind or the hard problem (and, as the [Gettierists](#) will point out, it's not really "knowing" either!). It's just back to the easy problem of doing. What makes it hard is not just that that "thought" is going on inside your head, but that it is going on in your mind, which is what makes it mental (another weasel-word).

Hence (by my lights) the only thing left to invoke to justify calling such a thing mental, and hence privileged, is that it feels like something to think it, and you are feeling that thing, and you're the only one in the position to attest to that fact.

Yet the fact you are attesting to here (as an eye-witness) is the fact that it *doesn't* feel like anything at all to think something! And that's why I think I am entitled to ask: Well then what is a thought, and how do you know you are thinking it? In what does your "consciousness of" it consist, if not that it feels like something to think it? (How can you be an eye-witness if your testimony is that you didn't see a thing?)

[Galen Strawson](#) has invoked the less weaselly (but nonetheless vague and hence still somewhat weaselly) notion of "experiencing" something, as opposed to "feeling it." But that just raises the same question: What is an unfelt "experience" (an experience it doesn't feel like anything to have)? Galen invokes "experiential character" or "phenomenal quality" etc. But (to my ears) that's either more weasel-words or just euphemisms because — for some reason I really can't fathom — one can't quite bring oneself to call a spade a spade. To be conscious of something, to experience something, to sense something, to think something, to "access" something — all of those are simply easy, toaster-like "information"-processing functions ("information" can be yet another weasel-word, if used for anything other than data, bits) — *except if they are felt*, in which case all functional, causal bets are off, and we are smack in the middle of the hard problem: why and how do feel? ([Ned Block's](#) unfortunate distinction between "phenomenal consciousness" and "access consciousness" is incoherent precisely because unfelt "access" is precisely what toasters have — hence "access" too is of the family *Mustelidae*, and the PC/AC distinction is bared as the attempt to distinguish felt feelings vs. unfelt "feelings" ...)

Amen — but with no illusions of having over-ridden anyone's (felt) privilege to insist that they do have unfelt [imageless?] thoughts — thoughts of which they are in some (unspecified) sense "conscious" even though it does not feel like anything at all to be conscious of them...

[Anil Seth](#)

[April 19th, 2011 at 6:30 pm](#)

Hello Stevan,

To begin with it's useful to recognize that your concept of a 'Turing biorobot' does improve on the unhelpful notion of a Zombie because you sensibly allow your biorobot to feel (experience, be aware (of), have consciousness even). As you put it, the challenge then becomes, How could a putative Turing biorobot help explain the how and the why of (calling a spade a spade) consciousness? I'll respond by recalling the history of the explanation of life — perhaps a well worn strategy but one worth rehearsing here. Once upon a time it seemed equally challenging to account for the seemingly ineffable property of life in terms of the doings of biological mechanisms, hence the powerful appeal to an *essence vital*. Nowadays we are relatively comfortable with the idea of life as a constellation concept which is mechanistically cashed out in a variety of processes (homeostasis, autopoiesis, metabolism, reproduction, etc.) no single one of which is necessary or sufficient. The 'hard' problem of consciousness may turn out much the same way. First, it is asking too much to explain why consciousness exists as an *explanandum*. Physicists have explained much about the universe without having to account for it's a priori existence (though the traffic from explananda to explanans testifies progress, as just now in the story of mass and the elusive Higgs boson). Second, at the level of 'feeling', consciousness can also be described as a constellation concept, consisting of (for example) a first-person-perspective integrating allocentric and egocentric components, a balance of differentiation and integration, volition,

agency, and the like. The right kind of doing explanation is then that which *accounts for* these properties or consciousness. For instance, neural dynamics that are simultaneously highly integrated and highly differentiated not only *correlate with* but also *account for* a key property of consciousness (that every conscious scene is one among many possible conscious scenes, and that every conscious scene is experienced as a unity). We might call these mappings ‘explanatory correlates of consciousness’, striking a contrast with the comparatively orthodox notion of ‘neural correlates of consciousness’, or ‘NCCs’, à la Crick and Koch. The still open question is then whether a comprehensive set of compelling explanatory correlates will dissolve the hard problem just as modern biology dissolved *essence vital*; once we can explain its properties apart from its very existence then the latter problem may be less troublesome. So let’s try to build a Turing Biorobot and see how it works.

Finally, the use of ‘feeling’ to mark out the terrain of the hard problem is dangerous because it reinforces a false distinction between cognition and affect with respect to consciousness. But as you rightly point out, thoughts – to the extent they can be divorced from affective states – are felt in the sense that they consist in or are accompanied by conscious contents. Conscious contents which convey meaning precisely because they are conscious.

Stevan Harnad

April 20th, 2011 at 5:04 pm

VITALISM, ANIMISM AND FEELING (Reply to Anil Seth)

Life Force: Many have responded to doubts about the possibility of explaining how and why organisms feel with doubts about the possibility of explaining how and why organisms are alive (suggesting that life is a unique, mysterious, nonphysical “life force”). Modern molecular biology has shown that the doubts about explaining life were unfounded: how and why some things are alive is fully explained, and there is no need for an additional, inexplicable life force. To have thought otherwise was simply a failure of our imaginations, and perhaps the same is true in the case of explaining how and why some organisms feel.

There is an interesting and revealing reason why this hopeful analogy is invalid: Life was always a bundle of objective properties (a form of “doing,” in the plain-talk gerunds of my little paper): Life *is* as life *does* (and can do): e.g., move, grow, eat, survive, replicate, etc., and just about all those doings have since been fully explained. Nor was there anything about those properties that was ever inexplicable “in principle”: No vitalist could have stated what it was about living that was inexplicable in principle, nor why. It was simply unexplained, hence a mystery.

But in the case of the mind/body problem, we can say (with Cartesian certainty) exactly what it is about the mind that is inexplicable (namely, feeling) and we can also say why it is inexplicable: (1) there is no evidence whatsoever of a “psychokinetic” mental force, (2) doing (the “easy problem”) is fully explainable without feeling, and (3) hence there is no causal room left for feeling in any explanation, yet (4) each of us knows full well that feeling exists. Hence the “hard” problem in the case of feeling is not based on the assumption there must be a non-physical “mental force” but on the fact that feeling is real yet causally superfluous.

(My guess is that what vitalists really had in mind all along, without realizing it, when they doubted that life could be explained physically and thought it required a nonphysical “vital force” was in fact feeling! It was actually the mentalism [animism] latent in the vitalism that was driving vitalists’ intuitions. Well, they were wrong about life, and they could never really have given any explicit reason in the first place for their doubts that living could not prove fully explainable physically and functionally (i.e., in terms of “doing”), just like everything else in nature. But with feeling we do have the explicit reason and it is not invalidated by the analogy with living. The important thing to bear in mind, however, is that the “hard” problem is not necessarily an “ontic” one. I don’t doubt that the brain causes feeling: I doubt that we can explain how or why, the way we can explain how the brain causes doing. The problem is with explaining the causal role of feeling, because that causal role cannot really be what it feels like it is.)

Feeling as Explanandum: I don’t think there’s any way for us to wriggle out of the need to explain how and why we feel (rather than just do) by an analogy with the fact that, say, physicists do not ask why the fundamental forces exist: We do not to ask why there is gravity; we just need to show how it can do what it does. If psychokinesis had been a fifth fundamental force, we could have accepted that as given too, and just explained how it can do what it does. But there’s no psychokinetic force. So it remains to explain how and why we feel, even though feeling, causally superfluous for all our doings, is undeniably there!

“Explanatory Correlates of Consciousness”: The only property of consciousness (a weasel-word) that is hard to explain is how and why anything is felt. The rest is “easy”: just an explanation of what our brains and bodies and mouths can do. Yes, the Turing Biorobot is the target to aim at. But it will not solve the hard problem — it

won't even touch it.

Conscious States = Felt States: Although the English word "feel" happens to be used mostly to refer to emotion and to touch, not only does it also *feel like something* to see, hear, taste, move and smell (in French "je sens" refers to emotion and to smell), but it also feels like something to think, believe, want, will, and understand. It is not "intentionality" (a weasel-word for the fact that mental [another weasel-word] states are "about" something or other) that is the mark of the mental, but the fact that mental states are felt states. Hence, apart from their "correlated" doings, "conscious states... convey meaning precisely because they are" felt. (The problem, as ever, is explaining how and why they are felt — rather than just done.)

David Rosenthal

April 19th, 2011 at 7:30 pm

Is there a hard problem about feeling? I think not. I think that the way to see this is by doing a bit of divide-and-conquer.

There are two aspects to conscious feelings. One is what they all have in common; they're all conscious. The other is their specific qualitative character, in respect of which feelings and conscious sensations and perceptions differ in type. The apparent intuition that there is a hard problem—a difficulty in explaining how conscious qualitative states can be subserved by or perhaps even identical with neural states—always proceeds by taking these two factors to be inseparable. And the apparent intuition that they're inseparable is, in turn, generated by the idea that we know about feelings—conscious qualitative states—solely by the way they present themselves to consciousness.

That's why there even *seems* to be an other-minds problem—which Stevan Harnad mentions, but says relatively little about. There seems to be an other-minds problem if one takes for granted to we know about qualitative states at bottom only by how they present themselves to consciousness; how, then, can we be sure anybody else is in such states?

I speak of the *apparent* intuitions that there is a hard problem and that qualitative character is inseparable from consciousness because I think these views are not, as they're often claimed to be, commonsense, pretheoretic intuitions at all, but are rather creatures of a particular theory, now so widespread as to encourage the idea that it's just common sense. That theory is what I mentioned before—that we know about conscious qualitative states solely by the way they present themselves to consciousness.

That this is just a theory can be seen by noting that there is a theoretical alternative: that we know about conscious qualitative states, our own as well as others', by way of the role they play in perceiving. Even bodily sensations such as pains are perceptual in nature; they enable us to perceive, typically if not always correctly, damage to our bodies. And we individuate types of qualitative character, in our own cases and in those of others, by appeal to the kind of stimulus that typically occasions the qualitative state in question. That's clear in the case of ordinary perceiving, but it holds for body sensations as well; we individuate types of pains by location and by typical feels of stabbing, burning, throbbing, and so forth.

Professor Harnad in effect dismisses this view, by relegating perceiving to mere doing, instead of feeling. But consider the individuation of states according to perceptual role; that will include states with the mental qualities of color, sound, shape, pain of various types, and all the other mental qualities.

But what about consciousness? We know that perceptual states can occur with qualitative character but without being conscious. Nonconscious perceptions occur in subliminal perceptions, e.g., in masked priming, and in blindsight, and we distinguish among such nonconscious perceptions in respect of qualities such as color, pitch, shape, and so forth. So qualitative character can occur without being conscious. What is it, then, in virtue of which qualitative states are sometimes conscious?

If, as in masked priming or other forced-choice experiments, we have reason to believe somebody is in a qualitative state that the person is wholly unaware of being in, we regard that state as not being conscious. So a conscious state is one that an individual is aware of being in. I have argued elsewhere that such higher-order awareness is conferred by having a thought to the effect that one is in that state; but the particular way that higher-order awareness is implemented isn't important for present purposes.

We have then, the makings of an explanation of how neural processing subserves—and is perhaps identical with—conscious qualitative states. But conscious qualitative states—what Professor Harnad calls feelings—are not, as is typically thought today, indissoluble atoms; rather they consist of states with qualitative character, which in itself occurs independently of consciousness, together with the higher-order awareness that results in those qualitative states' being conscious.

Stevan Harnad

April 20th, 2011 at 5:06 pm

UNFELT FEELINGS AND HIGHER-ORDER THOUGHTS (Reply to David Rosenthal)

DR: *"There are two aspects to... feelings. One is what they all have in common; (1) they're all conscious. The other is (2) their specific qualitative character."*

I would say the same thing thus:

There are two aspects to feelings. One is that (1) they're all felt. The other is (2) what each feeling feels like.

DR: *"The apparent... hard problem [is] a difficulty in explaining how conscious qualitative states can be subserved by or perhaps even identical with neural states"*

For me, the hard problem is not this metaphysical one, but the epistemic problem of explaining how and why some neural states are felt states.

DR: *"the apparent intuition that [(1) and (2)] are inseparable is... generated by the idea that we know about feelings... solely by the way they present themselves to consciousness."*

We know that we feel (1), and what each feeling feels like (2), because we feel it.

DR: *"There seems to be an other-minds problem if.. we know about qualitative states... only by how they present themselves to consciousness; how... can we be sure anybody else is in such states?"*

We each know for sure that we feel because we each feel. How can we each know for sure that anyone else feels? (There are plenty of reliable ways to infer it: that's not what's at issue.)

DR: *"these views are not... commonsense, pretheoretic intuitions... but... [the] theory... that we know about conscious qualitative states solely by the way they present themselves to consciousness."*

We know for sure that we feel (1), and what each feeling feels like (2), because we feel it. (That does not sound very theoretical to me!)

We each know for sure that we feel because we each feel. We can reliably infer that (and what) others feel (we just can't know for sure). (Only the inferring sounds theoretical here.)

DR: *"there is a theoretical alternative: that we know about conscious qualitative states, our own as well as others', by way of the role they play in perceiving."*

I know (a) that I feel, and (b) what I feel, and (c) that others feel, and (d) what others feel because of the role (a) – (d) play in "perceiving"?

But is perceiving something I feel or something I do? If it's something I feel, this seems circular. If it's something I do — or something I'm able to do — then we're back to doing, and the easy problem; how / why any of it is felt remains unexplained. (And "role" doesn't help, because it is precisely explaining causality that is at issue here!)

DR: *"Even bodily sensations such as pains are perceptual...; they enable us to perceive...damage to our bodies. And we individuate types of qualitative character, in our own cases and in those of others, by appeal to the kind of stimulus that typically occasions the qualitative state in question... we individuate types of pains by location and by typical feels of stabbing, burning, throbbing, and so forth."*

All true: Bodily stimulation and damage are felt. And, in addition, our brains know what to do about them.

But how and why are they felt, rather than just *detected*, and *dealt with* (all of which is mere doing)?

(Perception is a weasel-word. It means both detecting and feeling. Why and how is detection felt, rather than just done?)

DR: *"Professor Harnad... dismisses this view, by relegating perceiving to mere doing, instead of feeling. But... the individuation of states according to perceptual role... will include states with... color, sound, shape, pain of various types, and all the other mental qualities."*

Doing is doing. If it is *felt* doing then it is "perception": But how and why are (some) doings felt?

Yes, different feelings feel different. But the hard problem is explaining how and why any of them are felt at all.

DR: *"But what about consciousness? We know that perceptual states can occur with qualitative character but without being conscious. Nonconscious perceptions occur in subliminal perceptions, e.g., in masked priming, and in blindsight, and we distinguish among such nonconscious perceptions in respect of qualities such as color, pitch, shape, and so forth. So qualitative character can occur without being conscious. What is it, then, in virtue of which qualitative states are sometimes conscious?"*

Unfelt properties are not "qualitative states." Qualitative states are felt states. The roundness of an apple is not

a qualitative state, neither for the apple, nor for the robot that detects the roundness, nor for my brain if it detects the roundedness but I don't feel anything. The hard problem is explaining how and why I do feel something, when I do, not how my brain does things unfeelingly. (Why doesn't it do all of it unfeelingly?)
DR: *"If, as in masked priming or other forced-choice experiments... somebody is in a qualitative state that the person is wholly unaware of being in, we regard that state as not being conscious. So a conscious state is one that an individual is aware of being in."*

A felt state is a state that it feels like something to be in. "Unfelt feelings" do not make sense (to me). The only thing the unconscious processing (sic) and blindsight data reveal is how big a puzzle it is that not all of our internal states — and the doings they subservise — are unfelt, rather than just these special cases.

DR: *"I have argued elsewhere that such higher-order awareness is conferred by having a thought to the effect that one is in that state; but the particular way that higher-order awareness is implemented isn't important for present purposes."*

We cannot go on to the (easy) problem of "higher-order awareness" until we have first solved the (hard) problem of awareness (feeling) itself.

DR: *"We have then, the makings of an explanation of how neural processing subserves—and is perhaps identical with—conscious qualitative states."*

I am afraid I have not yet discerned the slightest hint of an explanation of how or why some neural states are felt states.

DR: *"But conscious qualitative states—what Professor Harnad calls feelings—are not, as is typically thought today, indissoluble, atoms; rather they consist of states with qualitative character, which in itself occurs independently of consciousness, together with the higher-order awareness that results in those qualitative states' being conscious."*

I don't see how one can separate a feeling from the fact that it is felt. I don't believe in (or understand) "unfelt feelings." Unfelt states are both unfelt, and unproblematic (insofar as the "hard" problem is concerned).

And "higher-order awareness" — awareness of being aware, etc. — seems to me to be among the perks of being an organism than can do what we can do, as well as of the (unexplained) ability to feel what it feels like to do and be able to do some of those things.

Yes, not only does it feel like something to be in pain, or to see red, but it also feels like something to contemplate that "I feel, therefore I'm sure there's feeling, but I can't be sure whether others feel (or even exist), but I can be pretty sure because they look and act more or less the same way I do, and they are probably thinking and feeling the same about me... etc." And that's quite a cognitive accomplishment to be able to think and feel that.

But unless someone first explains how and why anything can feel anything at all, all that higher-order virtuosity pales in comparison with that one unsolved (hard) problem.

Joel Marks

April 19th, 2011 at 10:07 pm

Most provocative, Stevan, but I wonder about two things (which are related to each other). First, why did you omit a category of attitudes, or whatever rubric belief and desire would fall under, as something cognitive science needs to explain (or explain away)? Surely they have been a preoccupation. And you do mention believing and understanding and doubting elsewhere in your paper.

Second, what do you think about this example (a variation of which I offered during the discussion of Jesse Prinz's paper although to make a different point)? Suppose you strongly desired to live but believed you were about to die, and let's say rather awfully to boot, which is something else you have a strong desire, i.e., an aversion about. Now I am assuming that you count beliefs and desires as non-feelings, although you also say there is something it feels like to be in those states. So there could be zombies with the belief/desire configuration I just mentioned who nonetheless felt nothing. Yet I wonder what it is they would be missing? For example, would we feel it was any less important to try to prevent the zombie from being in this configuration than the person who also (?) was feeling what it was like?

I must say that I have become skeptical altogether about feelings, at least as belonging to a distinct realm of sensations. Thank you.

Stevan Harnad

April 20th, 2011 at 4:31 pm

DISPOSITIONS TO DO: FELT AND UNFELT (Reply to Joel Marks)

JM: *"why did you omit a category of attitudes, or whatever rubric belief and desire would fall under, as something cognitive science needs to explain...?... you do mention believing and understanding and doubting..."*

Yes, cognitive science needs to explain attitudes, dispositions and tendencies. They are all part of the "easy" problem: doing.

Believing, desiring, understanding and doubting, besides having an "easy" aspect (dispositions and capacities to do) are also felt. That is the "hard" problem: Why are they felt, rather than just done (i.e., acted upon)?

JM: *"I am assuming that you count beliefs and desires as non-feelings, although you also say there is something it feels like to be in those states."*

No, believing and desiring are felt — though some people (not me) speak (loosely) of unfelt tendencies as "beliefs": I think that just creates confusion between doing and feeling, things that are, respectively, easy and hard to explain.

JM: *"Suppose you strongly desired to live but believed you were about to die...rather awfully... So there could be zombies with [that] belief/desire... who nonetheless felt nothing. I wonder what it is they would be missing?"*

What Turing robots would be missing if they did not feel would be feeling. And in that case they wouldn't have beliefs of desires either: they would just be *behaving as if* they had beliefs or desires (doing). All they would really have would be capacities and dispositions to do. (But I actually believe that a Turing-scale robot *would* feel — though of course we have no way of knowing...)

JM: *"would we feel it was any less important to try to prevent the zombie from [desiring to live/believing it would die] than the person who...was feeling... it?"*

(1) There is no way to know whether a Turing robot feels.

(2) For me it's likely enough that it would feel (so I wouldn't kick one).

(3) If there could be a guarantee from a deity that the Turing robot was a "Zombie" — as feelingless as a toaster — I suppose it would not matter if you kicked it (except for the wantonness of kicking even a statue). But there are no reliable deities from whom you can know that, so no way to know whether there can be Zombies.

(4) So the question is moot, since the answer depends entirely on unknowables.

JM: *"I must say that I have become skeptical altogether about feelings, at least as belonging to a distinct realm of sensations"* Skeptical that you feel? (That doesn't make sense to me.)

Skeptical about whether feelings are distinct from sensations? Anything felt is felt. If stimuli (of any kind — optical, acoustic, mechanical, chemical) are felt, they are sensations; if they are merely detected by your brain, but unfelt, then they are not sensations but merely receptor activity, peripheral or central.

In addition, there are other kinds of feelings, besides sensations: emotional, conational and cognitive feelings. Any state that it *feels like something* to be in.

David Rosenthal

April 20th, 2011 at 11:05 am

1.

DR: "There are two aspects to... feelings. One is what they all have in common; (1) they're all conscious. The other is (2) their specific qualitative character."

SH: "I would say the same thing thus:

"There are two aspects to feelings. One is that (1) they're all felt. The other is (2) what each feeling feels like."

I don't think that those two things are at all the same thing. We need an argument—not just an assertion—that the states we call feelings can't occur without being conscious (i.e., without being felt). If somebody doesn't want to apply the term 'feeling' to the ones that aren't conscious, fine; but I'm maintaining that that the very same type of state occurs sometimes as conscious qualitative states and sometimes not consciously.

That's a position in that space here, and it needs to be addressed, not just denied.

SH: "Unfelt properties are not 'qualitative states.' Qualitative states are felt states."

That's just the deal of my view.

SH:

2.

SH: "But is perceiving something I feel or something I do? If it's something I feel, this seems circular. If it's something I do — or something I'm able to do — then we're back to doing, and the easy problem; how/why any of it is felt remains unexplained. (And "role" doesn't help, because it is precisely explaining causality that is at issue

here!)”

I don't understand what it would be for perceiving to be something one feels. For one thing, that begs the question about whether perceiving can occur without being conscious. For another, it seems plain that it can occur without being conscious, as evidenced by subliminal perceiving, and so forth.

SH: “(Perception is a weasel-word. It means both detecting and feeling. Why and how is detection felt, rather than just done?)”

Well, I don't know that it's a weasel word—though I agree that it means both. In the nonconscious case it's (mere) detecting; in the conscious case, it's *conscious* detecting, i.e., feeling.

3.

SH: “Doing is doing. If it is felt doing then it is 'perception': But how and why are (some) doings felt? ... But the hard problem is explaining how and why any of them are felt at all. ... The hard problem is explaining how and why I do feel something, when I do, not how my brain does things unfeelingly. (Why doesn't it do all of it unfeelingly?)”

There are two issues here. One is to explain why some qualitative states—some perceivings—come to be conscious; why don't all remain subliminal? I think that's a difficult question, which I address in several publications (e.g., CONSCIOUSNESS AND MIND), but I won't address here.

I don't think, however, that it's reasonable to assume that everything has a utility or function, so that it can't be the case that at least some of the utility or functionality of perceiving occurs consciously. Not everything that occurs in an organism is useful for the organism. But that's for another day.

4.

But the other question is what is it in virtue of which conscious qualitative states differ from qualitative states that aren't conscious. I addressed that earlier:

DR: “If, as in masked priming or other forced-choice experiments... somebody is in a qualitative state that the person is wholly unaware of being in, we regard that state as not being conscious. So a conscious state is one that an individual is aware of being in.”

The only way to deny that, so far as I can tell, is by denying that any qualitative states occur without being conscious. But subliminal cases give us good reason to think that they do.

SH: “We cannot go on to the (easy) problem of 'higher-order awareness' until we have first solved the (hard) problem of awareness (feeling) itself.”

That begs the question against the higher-order theory—and the occurrence of nonconscious qualitative states.

Stevan Harnad

April 21st, 2011 at 3:42 pm

“UNCONSCIOUS FEELING” VS. “UNFELT CONSCIOUSNESS”: DETECTING THE DIFFERENCE (Reply to David Rosenthal-2)

DR: “We need an argument – not just an assertion – that the states we call feelings can't occur without being conscious (i.e., without being felt).”

I agree that an argument is needed, but I'm not sure whether it's the affirmer or the denier who needs to make the argument!

First, surely no argument is needed for the tautological assertion that feelings and felt states have to be felt: (having) an unfelt feeling or (being in) an unfelt felt-state is a contradiction.

The more substantive assertion is that all unfelt states are unconscious states and all felt states are conscious states (i.e., feeling = consciousness). And its denial would be either that (1) no, there can be states that are unfelt, yet conscious, or (2) no, there can be states that are unconscious yet felt (or both).

I would say the burden, then, is on the denier, either (1) to give examples of unfelt states that are nevertheless conscious states — and to explain in what sense they are conscious states if it does not feel like anything to be in those states — or (2) to give examples of unconscious states that are nevertheless felt states — and to explain who/ what is feeling them if it is not the conscious subject — (or both).

(It will not do to reply, for (1), that the subject in the conscious state in question is indeed awake and feeling something, but not feeling what it feels like to be in that state. That makes no sense either. Nor does it make sense to reply, for (2), that the feeling is being felt by someone/ something other than the conscious subject.)

What you have in mind, David, I know, is things like “unconscious perception” and blindsight. But what's meant by “unconscious perception” is that the subject can be blind to something that is perceived by someone else.

meant by unconscious perception is that the subject somehow behaves as if he had seen something, even though he is not conscious of having seen it. For example, he may say he did not see a red object presented with a masking stimulus, and yet respond more quickly to the word "dead" than to "glue" immediately afterward (and vice versa if the masked object was blue).

Well there's no denying that the subject's brain detected the masked, unseen red under those conditions, and that that influenced what the subject did next. But the fact is that he did not see the red, even though his brain detected it. Indeed, that is why psychologists call this *unconscious* "perception." That's loose talk. (It should have been called unconscious *detection*.) But in any case, it is *unconscious*. So it does not qualify as an instance of something that is conscious yet unfelt.

But, by the same token, it also does not qualify as an instance of something that is unconscious yet felt: Felt by whom, if not by the conscious subject? You don't have to feel a thing in order to "detect": Thermostats, robots and other sensors do it all the time. Detecting is something you do, not something you feel.

As for blindsight, some of it may be based on feeling after all, just not on visual feeling but other sensory but nonvisual (e.g. kinesthetic) feelings (for example, feeling where one's eyes are moving, a doing that is under the control of one's intact but involuntary and visually unconscious subcortical eye-movement system).

But some blindsight may indeed be based on unconscious detection — which is why the patient (who really can't see) has to be encouraged to point to the object even though he says he can't see it. It's rather like automatic writing or speaking in tongues, and it is surprising and somewhat disturbing to the patient, who will try to rationalize (confabulate) when he is told and shown that he keeps pointing correctly in the direction of the object even though he says he can't see a thing.

But this too is neither unfelt consciousness nor unconscious feeling: If the subject is not conscious of seeing anything, then that means he is not feeling what it feels like to see. And if he's not feeling it, neither is anything or anyone else in his head feeling it (otherwise we have more than one mind/body problem to deal with!). If he can nevertheless identify the object before his eyes, then this is unfelt doing, not unconscious feeling.

All these findings simply compound the hard problem: If we don't really have to see anything in order to detect stimuli presented to our eyes, then why does it feel like something to see (most of the time)?

Ditto for any other sense modality, and any other thing we are able to do: Why does it feel like something to do, and be able to do all those things?

And this "why" is not a teleological "why": It's a functional why. It's quite natural, if you have a causal mechanism, consisting of a bunch of widgets, to ask: What about this widget: What's it doing? What causal role is it playing? What do you need it for?

Normally, there are answers to such questions (eventually).

But not in the case of feeling. And that's why explaining how and why we feel is a "hard" problem, unlike explaining how and why we do, and can do, what we do. Explanations of doing manage just fine, to all intents and purposes, without ever having to mention feeling (except to say it's present but seems to have no causal function).

DR: "If somebody doesn't want to apply the term 'feeling' to the [states] that aren't conscious, fine; but I'm maintaining that the very same type of state occurs sometimes as conscious qualitative states and sometimes not consciously."

Unfelt detection cannot be *the very same state* as felt detection otherwise we really would have a metaphysical problem! What you must mean, David, is that the two kinds of states are similar in some respects. That may well be true. But the object of interest is precisely the respect in which the states differ: one is felt and the other is not. What functional difference does feeling make (i.e., why are some states felt states?), and how?

And, to repeat, blueness is a quality (i.e., a property — otherwise "quality" is a weasel-word smuggling in "qualia," another weasel-word which just means feelings). Blueness is a quality that a conscious seeing subject can feel, by feeling what it's like to see blue. One can call that a "qualitative state" if one likes (and one likes multiplying synonyms!). But just saying that it feels like something to see blue — and that to feel that something is to be in a felt state — seems to say all that needs to be said.

To detect blue without feeling that something it feels-like is to detect a quality (i.e., a property, not a "quale," which is necessarily a felt quality), to be sure, but it's not to be in "qualitative state" — unless a color-detecting sensor is in a qualitative state when it detects blue.

To insist on calling the detection of a quality "being in a qualitative state" sounds as if what you are wanting to invoke is unconscious feelings ("unconscious qualia"). But then one can't help asking the spooky question: Well then who on earth or what on earth is *feeling* those feelings. if it isn't the conscious subject?

There's certainly no need to invoke any spooks in me that are feeling feelings I don't feel when I am a subject in an unconscious perception experiment, since all that's needed is unconscious processing and unconscious detection, as in a robot or a tea-pot (being heated). A robot could easily simulate masked lexical priming with optical input and word-probability estimation. But no one would want to argue that either the robot or any part of it was feeling a thing, in detecting and naming red, and its knock-on effects on the probability of finding a word rhyming with "dead"...

DR: "[Your saying] "Unfelt properties are not 'qualitative states.' Qualitative states are felt states'" [is] just the denial of my view."

It depends on what is meant by "qualitative states." If the robot detecting and naming briefly presented red objects — and subsequently more likely to pick a word that rhymes with "dead" — is an instance of a "qualitative state," that's fine, but then we're clearly talking about the easy problem of doing (responding to optical input, processing words) and not the hard problem of feeling. Neither feeling nor consciousness (if there's any yet-to-be-announced distinction that can be made between them) plays any part in these same doings in today's robots.

(All bets are off with Turing-Test-scale robots — but even if they do feel — and I for one believe that Turing robots would feel — we still have to solve the problem of explaining how and why they do feel...)

DR: "I don't understand what it would be for perceiving to be something one feels. For one thing, that begs the question about whether perceiving can occur without being conscious. For another, it seems plain that it can occur without being conscious, as evidenced by subliminal perceiving, and so forth."

If it is unfelt, "perceiving" is just detecting (and responding). We know that countless unfeeling, unconscious devices can do detecting (and responding) without feeling. Hence the question is not whether detection can occur without feeling: it's how and why some detecting (namely, perceiving) is felt.

The burden of showing that one can make something coherent and substantive out of the putative difference between felt detection and conscious detection is, I have suggested, on the one who wishes to deny that they are one and the same thing. (That's why "perception" is a weasel-word here, smuggling in the intuition of felt qualities while at the same time denying that anyone is conscious of them.)

So subliminal "perceiving," if unfelt, is not perceiving at all, but just detecting.

DR: "Well, I don't know that ["perceiving" is] a weasel word – though I agree that it means both [detecting and feeling]. In the nonconscious case it's (mere) detecting; in the conscious case, it's conscious detecting, i.e., feeling."

Agreed!

But that does make it seem as if "feeling" and "consciousness" are pretty much of a muchness after all. And that whatever is or can be done via unfeeling/unconscious detection is unproblematic (or, rather, the "easy" problem), and what remains, altogether unsolved and untouched, is our "hard" problem of how and why some detection is felt/conscious...

DR: "There are two issues here. One is to explain why some qualitative states – some perceivings – come to be conscious; why don't all remain subliminal? I think that's a difficult question, which I address in several publications (e.g., Consciousness and Mind), but I won't address here."

Perhaps, David, in your next posting you could sketch the explanation, since that is the very question (for you "difficult," for others, "hard") that we are discussing here. If we are agreed that "unconscious" = "unfelt" = "subliminal" states are all, alike, the "easy" ones to explain, whereas the "felt" = "conscious" = "supraliminal" states are the "hard" ones to explain (and only weasel-words like "qualitative states" and "perception" have been preventing us from realizing that), then it's clearly how you address the hard problem (of explaining how and why some states are felt/conscious) that would be of the greatest interest here.

DR: "I don't think, however, that it's reasonable to assume that everything has a utility or function, so that it can't be the case that at least some of the utility or functionality of perceiving occurs consciously. Not everything that occurs in an

organism is useful for the organism. But that's for another day."

I couldn't quite follow the middle clause (beginning "so that it can't be the case"), but it sounds as if you are suggesting that there may be no functional/causal explanation for why some doing and doing-capacity is felt. I'm not sure what would be more dissatisfying: that there is no way to explain how and why some functional states are felt, or that some functional states are felt for no reason at all! The first would be a huge, perplexing fact, being doomed to remain unexplained; the other would be a huge, perplexing fact being a mere accident.

DR: "[Your saying] "We cannot go on to the (easy) problem of 'higher-order awareness' until we have first solved the (hard) problem of awareness (feeling) itself'" begs the question against the higher-order theory – and the occurrence of nonconscious qualitative states."

I think we've agreed that calling unfelt/unconscious states "qualitative" is merely a terminological issue. But, on the face of it, bootstrapping to higher-order awareness without first having accounted for awareness (feeling) itself seems rather like the many proofs — before the proof of Fermat's Last Theorem — of the higher-order theorems that would follow from Fermat's Last Theorem, *if Fermat's Last Theorem were true*. Maths allows these contingent necessary truths — following necessarily from unproved premises — because maths really is just the study of the necessary formal consequences of various assumptions (e.g., axioms).

But here we are not talking about deductive proofs. We are talking about empirical data and (in the case of cognitive science, which is really just a branch reverse bioengineering) the causal mechanisms that generate those empirical data.

So it seems to me that a theory of higher-order consciousness is hanging from a skyhook if it has not first explained consciousness (feeling) itself.

Shimon Edelman

April 20th, 2011 at 8:01 pm

As far as I can see, Stevan's challenge — to explain *how and why we feel rather than just do* — can only be met partly. And yet, the explanatory situation with regard to this question is much more benign than what a casual observer of the "hard problem" debates in philosophy might conclude. Indeed, I believe that Stevan is precisely right in denying the central tacit premise behind the challenge as phrased above: that feelings are somehow distinct from and independent of doings. Let me explain.

If one insists on addressing the question as posed, the first order of business should be to understand feelings (insofar as understanding doings is unproblematic). Now, the *how* part of the question is relatively easy, if one accepts that minds are what brains do (or, more generally, what systems that are brain-like in relevant respects do). I shall comment on it a bit later, but let me first turn to the *why* part.

On the one hand, there seems to be no principled way of explaining *why* a proverbial ripe tomato feels **this** way to me when I look at it, namely, red (as opposed to sweet, or fragrant, or squishy). Likewise, there is no principled way of explaining *why* it feels **that** way to me when I handle it, namely, squishy (as opposed to red, etc.) Asking *why* any quale feels the way it does amounts to a category mistake (to use an expression introduced by Ryle for a slightly different purpose), and so the only answer this question deserves is "Because."

On the other hand, as Stevan notes (and as certain other scientists and philosophers noted in the past, from Dan Dennett and Austen Clark all the way back to Karl Lashley and William James), there definitely is a principled way to explain *why* two shades of red feel **this different** to me. On the behavioral level, the domain of enquiry within which relevant explanations are constructed is psychophysics (as Stevan's discussion of just noticeable differences in this context suggests). In particular, Clark (1985) points out that qualia "*enable one to discern similarities and differences: they engage discriminations.*" With a little daring, this approach can be used to ground phenomenology in science. The conceptual leap that makes such grounding possible is akin to the explanatory move that is inherent in the Church-Turing Thesis. The CTT posits the equivalence of Turing computation, which is a formal concept, and effective computation, which is merely an intuitive one. Just like the CTT, the postulate equating qualia (feelings) with discriminations (doings) cannot be proved, yet if it gets things right, many explanatory benefits may ensue (including the demystification of meanings).

For such a move to work, the formal side of the equation (which speaks to the *how* part of Stevan's question) must be properly taken care of. A [paper](#) by Tomer Fekete and myself that suggests how to do so within a formal theoretical

framework of computation in dynamical systems, titled "Towards a computational theory of experience", is now in press in *Consciousness and Cognition*. Briefly, it equates the phenomenal experience of a mind — its *feeling* — with the dynamically unfolding trajectory that the collective state of the mind's brain — its *doing* — traces through an intrinsically structured space of possible trajectories (which defines the range of discernments or qualia that the mind in question can experience). You are invited to take the leap and join the exploration party on the other side.

Stevan Harnad

April 21st, 2011 at 11:23 am

DOING THE DOABLE — BUT IS "JUST BECAUSE" AN ANSWER? (Reply to Shimon Edelman)

A serious cognitive scientist ignores the rich and original work of Shimon Edelman at his peril. A master at relating perceptual differences to linguistic differences, if anyone is likely to put the "heterophenomenology" of Dan Dennett (another formidable thinker, in whatever JNDs one might differ from his views!) on a solid psychophysical and computational footing, it is Shimon.

But does the computational (or dynamical) explanation of each and every JND we can discriminate (all of which is *doing*) explain how or why those doings are *felt*?

(And if it does not, and "Because" is the only answer we can ever get to this "hard" question, does that mean it was unreasonable to have asked the question at all? I think this would be to paper over a fundamental explanatory crack — probably our most fundamental one. The "hard" problem may well be insoluble — but surely that does not mean it is trivial, or a non-problem, or that it was some sort of "category mistake" to have asked!)

SE: "Stevan is... right in denying the central... premise... that feelings are somehow distinct from and independent of doings."

Feelings and doings seem to be tightly *correlated*: that's undeniable. But it's the *causation* (and causal function) that's at issue here.

(And one can have reservations about feeling/doing commensurability too, for the psychophysical correlation is really only a doing/doing correlation: input/output. Inquiring more deeply into the "quality" of feelings, and their "resemblance" to things in the world, runs into Wittgensteinian private-language indeterminacy problems: what's the common metric? and what's the error-detector?)

SE: "the how part of the question is relatively easy, if one accepts that minds are what brains [and brain-like systems]... do"

Doing is what brains do. How and why they generate feelings — how and why it feels like something to do and to be able to do all that doing — is another matter.

But let me stress that the "why" in the "how and why" question is not an idle teleological query: It is a functional query, which means a causal query. If there are various functional components that generate our doing power, it seems reasonable to ask of each of them: "What causal role do they play in the successful outcome? What do they enable us to do that could not be done without them? What would be functionally missing or malfunctioning without them?"

In other words, the "why" in the "how and why" is just a call for a clear account of the specific causal contribution of feelings to the successful generation of our doing power, lest we simply take it for granted and forget that there's a huge elephant in the room whose presence still calls for an explanation in an account of doing that looks for all the world as if it would be equally compatible with the presence or the absence of feelings.

SE: "there seems to be no... way of explaining why a... tomato feels this way... when I look at it... red [and] that way to me when I handle it... squishy... Asking why any quale [feeling] feels the way it does amounts to a category mistake... [and] deserves... only [the] answer... 'Because'"

But the hard problem is not that of explaining how or why something feels this way rather than that way, but explaining how and why it feels like anything at all.

A category mistake is to ask whether an apple is true ("it's not true? well then it's false?"). There's no category error in asking how and why we feel rather than just do.

And if the answer is just “Because,” it’s not the impatient “Because” that questions like “why is there something rather than nothing?” or even “why is there gravity?” deserve. We are squarely in the world of doing, and its functional explanation. And there is a prominent property that is undeniably present but does not seem to have any causal role (despite the fact that, ironically, it feels causal — although that’s not the reason its presence calls for an explanation).

Waving that away with “Because” and “category error” is rather too quick...

SE: “there definitely is a... way to explain why two shades of red feel this different to me... psychophysics [JNDs]”

There definitely is a way to explain how and why we can discriminate everything we can discriminate — and manipulate and categorize and name and describe.

But those are all doings and doing capacities. How and why are they *felt* doings and doing capacities, rather than just “done” doings and doing capacities?

(The one making the category error here seems to be Shimon!)

SE: “The conceptual leap that makes such grounding possible is akin to the explanatory move that is inherent in the... Church-Turing Thesis [CTT]...the equivalence of Turing computation... a formal concept, and effective computation... an intuitive one.”

This is a very clever analogy — between, on the one hand, “capturing” intuitive computation with Turing computation, and, on the other, “capturing” peoples’ feelings with Turing models of doing — but it unfortunately cannot do the trick:

First, the fact that Turing computation merely “captures” mathematicians’ intuitions of what they mean by computation rather than proving that they are correct is the reason CTT is a *thesis* and not a *theorem*.

Mathematicians have other kinds of intuitions too — such as the Goldbach Conjecture, but those are theorems, subject to proof (such as in the recent proof of Fermat’s Last Theorem). Any thesis or conjecture can be invalidated by a single counter-example. But it takes a proof to show that it is true. The reason CTT cannot be proved true is that except when it is explicitly formalized, it is just a feeling! (Bertrand Russell, drawing on an example from William James, famously reminded mathematicians how feeling can be an unreliable guide too: “The smell of petroleum pervades throughout.”)

Now in science and engineering, we are not looking for proof of the truth of theorems but for evidence of the truth of theories. And evidence does not just mean gathering data that are compatible with and confirm the predictions of the theory. It means giving a causal explanation. This is clearest in engineering, where the way you test whether your theory successfully explains the way to get certain things done is to build a system (say, a vacuum cleaner) that tries to do those things according to the causal mechanism proposed by your theory, and show that the causal mechanism works (i.e., it can suck in dust).

Cognitive science is not basic science; it is more like reverse engineering, along lines similar to ordinary forward engineering. The only difference is that our vacuum cleaners grow on trees, so we have to try to reverse-engineer their doing-capacities and then test whether they have the causal power to generate out doings. That’s Turing’s method.

Now once we have a causal theory that is able to generate all of our doing power, we have causally explained doing (the “easy” problem). But have we “captured” feeling, the way the CTT has provisionally “captured” mathematicians’ intuitions about what computation is?

The answer is already apparent with CTT — which is, in a sense, also a cognitive theory of computation: a theory not only of what computation is, but of how computation is implemented in the brain. Computation, however, is *doing*, not feeling! So although mathematicians may have feelings about computation — just as we have feelings about tomatoes — it is not feelings that Turing computation implements but computations (doings). And (except for the blinkered believers in the computational theory of mind), feelings are not computations.

So the only sense in which a successful Turing theory of doing “captures” feeling is that it generates (and explains, causally) the doings that are correlated with feeling. It does not explain how or why those doings are felt or depend causally on feeling. The “conceptual leap” — to the conclusion that the successful explanation of doing explains feeling — is just as wrong in the case of Turing robotics as the notion that Turing computation (CTT) has explained mathematicians’ *feelings* about what computation is. Turing computation (provisionally)

captures what mathematicians are doing when they compute. There has not yet been a counterexample; maybe there never will be. That is unproven. Explaining how / why it feels like something for mathematicians to compute — and how / why it feels like something for mathematicians to think about computation and about what is and is not computation — is a gap not bridged by the “conceptual leap.”

SE: “Just like the CTT,... equating... feelings with discriminations (doings) cannot be proved... [but] many explanatory benefits may ensue (including the demystification of meanings).”

Many explanatory benefits ensue from explaining doing (i.e., solving the “easy problem”); that is uncontested. Shimon’s work especially will help integrate perceptual capacity with linguistic capacity. But that’s all on the doing side of the ledger.

SE: “Towards a computational theory of experience... equates... feeling – with the dynamically unfolding trajectory that the collective state of the... brain – its doing – traces through an intrinsically structured space of possible trajectories (which defines the range of [differences]... the mind... can [feel].”

This may define the range of differences that the brain can discriminate (do). But the fact that it’s felt alas remains untouched.

Shimon Edelman

April 22nd, 2011 at 9:42 am

Stevan writes “And (except for the blinkered believers in the computational theory of mind), feelings are not computations”, but does not pause to defend this claim. In the *Consciousness and Cognition* paper that I referred to in my first set of comments, Tomer Fekete and I argue that feelings in fact are computations, albeit not Turing computations — which is why Stevan’s jab at the blinkered Fodorians and his discussion of Turing computers prompted by my mention of the CTT are both beside the point. I cannot do justice here to the arguments laid out in our 22 kiloword paper, so let me just mention that it does take on all the issues raised in Stevan’s piece, including the causal role of feelings (which stems from their close relationship to discernments (JNDs) and therefore to the conceptual structure of the mind), while avoiding the panpsychism implied by computational “models” that are underconstrained by the intrinsic dynamics of the computational substrate. What we offer is an attempt at a principled and tightly constrained explanatory reduction of feelings to doings. Over and above the detailed arguments and the data that support them there is, however, the reductive leap that I alluded to earlier: feelings **are** doings in the sense that we discuss. Would you tell a physicist who offers you a theory of electrodynamics “Yes, I understand what electrons *do* with their charge, but what is charge and why do they have it?”? As Giulio Tononi noted (in connection with his Information Integration Theory, which I actually don’t quite buy), experience has the same ontological status as charge. In fact, its status is even more primary, given that experience is (as Ernst Mach and William James would agree) prior to everything else in the Universe. Yes, Stevan, cognitive science is a basic science :-)

Stevan Harnad

April 22nd, 2011 at 3:51 pm

INTERNAL/EXTERNAL ISOMORPHISM, DISCRIMINATION AND FEELING (Reply to Shimon Edelman-2)

SE: “Stevan... does not... defend [his] claim [that] ‘feelings are not computations... (except for the blinkered believers in the computational theory of mind).’ I argue that feelings in fact are computations, albeit not Turing computations...”

In his paper, Shimon makes it clear that by “computations” he does not just mean the Turing computations referred to by the Church-Turing Thesis: “[E]very physical process instantiates a computation insofar as it progresses from state to state according to dynamics prescribed by the

laws of physics, that is, by systems of differential equations.

Hence what Shimon means by “feelings are computations” is just that they are (somehow) properties of dynamical systems (hardware) rather than just hardware-independent Turing computations (formal symbol systems).

That’s not computationalism (the metaphysical theory that felt states are [Turing] computational states); it’s physicalism (the metaphysical theory that felt states are physical [dynamical] states). Well, yes, we’re all physicalists rather than “dualists”; but that doesn’t help solve the hard problem — of explaining *how and why* some physical states are felt states. This is not a metaphysical question but a functional one.

I have not yet read the “22 kiloword” [Fekete & Edelman paper](#) in detail, but I think I’ve understood enough of it to try to explain why I think it misses the mark, insofar as the hard problem is concerned:

The goal is to explain how and why we feel. The intuition (largely a visual one) is that external objects are dynamical systems, with (static and) dynamical properties (like size, shape, color) that (1) we feel because (2) they are “represented” in our brain by another system — an internal dynamical system that mirrors (and can operate on) those dynamical properties, right down to the last JND (just-noticeable-difference).

Now the Fekete & Edelman model is not yet implemented. But if ever it is, it is very possible that it might help in generating some of our capacity to do what we can do. Let’s even suppose it can generate all of it, powering a Turing robot that can do anything and everything we can do, right down to the last JND.

And we know how it does it: It has an internal dynamical system that mirrors the properties of the external dynamical systems that we can see, hear, manipulate, name and describe. That solves the “easy” problem of doing.

Now what about the feeling? If the Turing robot views a round shape, it can do with it all the things we can do with round shapes, in virtue of its internal dynamical counterparts, including the minutest of sensory discriminations (though one wonders why internal representations are needed to do same/different judgments on externally presented pairs of round shapes of identical or minutely different size). In any case, the internal analogs may come in handy for tasks such as the [Shepard internal rotation task](#). (I say “internal” rather than “mental,” because “mental” would be a weasel-word here insofar as the question of feeling versus doing is concerned here.)

The internal representations of shape certainly mirror the shapes of the external objects, but do they mirror what it *feels like* to see round shapes? How? I mean, if we made a trivial toy robot that could only do same/different judgments on round shapes, or on rotated Shepard-shapes, would it be feeling anything, in virtue of its internal dynamics? Why not? Would scaling up its capacity closer and closer to ours eventually make it start feeling something? when? and how? and why?

So, no, although the idea of generating internal dynamical representations that are isomorphic to external objects is a natural intuition about how to go about building what it feels like to perceive the world into a brain or robot, all it really does is give the brain or robot a means of doing what it can do (including the minutest discrimination, all the way down to a JND). It’s an input/output isomorphism, not an input/feeling isomorphism. It is as unexplained as ever why anything should be felt at all, under any of these conditions. Why should it feel like something to discriminate? Discriminating is doing. All that’s needed is the power to do it.

And there’s also the question of commensurability: Internal and external shapes are commensurable; so are input shapes and output shapes. But what about the commensurability of external shapes and what it feels like to see them? They are commensurable only on condition that the internal analogs are indeed, somehow, felt, rather than just used to *do* something (like making same/different judgments for successive rotated and unrotated external shapes). But why are they felt?

So I would say that such internal analogs and their dynamics may very well cast some light on the easy problem of how the brain can do some of the things it can do, but that they leave completely untouched the hard problem of how and why it *feels like something* to do or be able to do what the brain can do.

SE: “the causal role of feelings... stems from their close relationship to discernments (JNDs) and therefore to the conceptual structure of the mind”

But how and why are discernments (JNDs) *felt*, rather than just *done*?

SE: “[Our model] avoid[s] the panpsychism implied by computational “models” that are underconstrained by the intrinsic dynamics of the computational substrate”

In other words, make sure that the internal/external isomorphism is tight enough and specific enough to avoid the conclusion that “any kind of organized matter [feels] to some extent.” Agreed. But it remains to explain how and why any kind of organized matter — whether or not isomorphic up to a JND — feels to any extent at all!

SE: “What we offer is an attempt at a principled and tightly constrained explanatory reduction of feelings to doings... a reductive leap [to the effect that] feelings are doings in the sense that we discuss.”

Shimon, I’m afraid the reductive leap doesn’t work for me! Doings are still doings, and it’s not at all clear how or why internal analog dynamics that mirror external dynamics in the service of discriminating or any other doing should be *felt* dynamics rather than just *done* dynamics.

SE: “[C]ognitive science is a basic science :-) ... [Feeling] has the same ontological status as charge... Would you tell a physicist who offers you a theory of electrodynamics ‘Yes, I understand what electrons do with their charge, but what is charge and why do they have it?’?”

No, I wouldn’t, because it’s evident that the question-asking must stop with the four basic forces of nature (electromagnetism, gravitation, the weak force and the strong force). But feeling (unless you are a panpsychist despite the complete absence of evidence for a psychokinetic force) is not one of the basic forces of nature. And cognitive science is not a basic science! :-)

So I’m inclined to repeat what I said in my first reply: If “Because!” is the only answer we can ever get to our “hard” question, does that mean it was unreasonable to have asked the question at all? I think this would be to paper over a fundamental explanatory crack — probably our most fundamental one. The “hard” problem may well be insoluble — but surely that does not mean it is trivial, or a non-problem, or that it was some sort of “category mistake” to have asked!

John Campbell

April 21st, 2011 at 7:03 pm

Stevan’s point is, ‘It feels like something to mean something’. Hence, problems of meaning shouldn’t be separated from ‘the hard problem of explaining how and why we feel’, which is ‘even more pervasive than sensory and emotional experience’. Stevan uses Searle’s Chinese Room to make his point vivid – discussing meaning in the absence of hard problems leaves one with ‘blind-semantics’.

I wonder whether ‘feeling’ is really the right notion to fasten onto here. The obvious problem with ‘blind-semantics’ as illustrated by the Chinese Room is that language-use is being described without giving any place to its relation to perceptual experience. The connection between meaning and the hard problems comes when we try to characterize the relation between meaning and our sensory awareness of our surroundings.

I agree that there is such a thing as ‘speaking with feeling’, or ‘feeling the full weight of what one is saying’, for example. But the fundamental point of contact with the hard problem has to do with sensory awareness. Once one has grasp of meaning, suitably hooked up to sensory experience, in an agent with some kind of emotional life, then as a *consequence* of that there will be such a thing as ‘it feeling like something to mean something’, but that’s an epiphenomenon.

Stevan Harnad

April 22nd, 2011 at 9:17 am

SENSORIMOTOR GROUNDING OF WORDS IS NECESSARY BUT NOT SUFFICIENT FOR MEANING
(Reply to John Campbell)

JC: “ I wonder whether ‘feeling’ is really the right notion to fasten onto here. The obvious problem with ‘blind-semantics’ as illustrated by [Searle’s] Chinese Room is that language-use is being described without giving any place to its relation to perceptual experience. The connection between meaning and the hard problems comes when we try to characterize the relation between meaning and our sensory awareness of our surroundings.”

Meaning = Sensorimotor Grounding + Semantic Interpretability + Feeling. Yes, computation (formal symbol manipulation) alone is not enough for meaning, even if it has a systematic semantic interpretation. This is the “symbol grounding problem” (one of the “easy” problems).

The solution to the symbol grounding problem is to ground the internal symbols (words) of a Turing robot in its autonomous sensorimotor capacity to detect, categorize, manipulate and describe the symbols’ external referents.

But although grounding is necessary for meaning, it is not sufficient. The other necessary component is feeling: It *feels like something* to mean something. If I say “The cat is on the mat,” I am not only generating a syntactically well-formed string of symbols — part of a symbol system that also allows me to systematically generate other symbol strings, such as “The cat is not on the mat” or “The mat is on the cat” or “The rat is on the mat,” etc. — all of which are systematically interpretable (by an *external* interpreter) as meaning what they mean in English. In addition to that semantic interpretability to an external interpreter, I am also able, autonomously, to detect and interact with cats and mats, and cats being on mats, etc., with my senses and body, and able to interact with them in a way that is systematically coherent with the way in which my symbol strings are interpretable to an external interpreter.

I have no idea whether there can be “Zombies” — Turing robots whose doings and doing-capacity are indistinguishable from our own but that do not feel — although I doubt it. (I happen to believe that anything that could do what a normal human being can do — indistinguishably from a human, to a human, for a lifetime — would feel.)

But my belief is irrelevant, because there’s no way of knowing whether or not a Turing robot (or biorobot) is a Zombie: no way of determining whether there can be Turing-scale grounding without feeling. Worse, either way there is no explanation of feeling: neither an explanation of how and why a grounded Turing robot feels, if it is not a Zombie, nor an explanation of how and why we feel and the Turing robot doesn’t, if it’s a Zombie.

But what is clear is what the difference would be: the presence or absence of feeling. And that is also the difference between meaning something or merely going through the motions.

JC: “ I agree that there is such a thing as ‘speaking with feeling’, or ‘feeling the full weight of what one is saying’ , for example. But the fundamental point of contact with the hard problem has to do with sensory awareness. Once one has grasp of meaning, suitably hooked up to sensory experience, in an agent with some kind of emotional life, then as a consequence of that there will be such a thing as ‘it feeling like something to mean something’, but that’s an epiphenomenon.”

It may well be that most of what it feels like to mean “the cat is on the mat” is what it feels like to recognize and to imagine cats, mats, and cats being on mats.

But the bottom line is still that to say (or think) and mean “the cat is on the mat” there has to be something it feels like to say (or think) and mean “the cat is on the mat” — and that *for someone to be saying (or thinking) and meaning “the cat is on the mat” they have to be feeling something like that*. Otherwise it’s still just “blind semantics,” even if it’s “suitably hooked up (grounded) in sensorimotor capacity.

(Sensory “experience,” by the way, would be an equivocal weasel-word, insofar as feeling is concerned: is it *felt* experience or just “done” experience, as in a toaster or a toy robot?)

So I’m definitely not speaking of “speaking with feeling” in the sense of emphasis, when I say there’s something it feels like to mean something (or understand something).

I mean that to mean “the cat is on the mat” (whether in speaking or just thinking) is not just to be able to generate word strings in a way that is semantically interpretable to an external interpreter, nor even to be able to interact with the cats and mats in a way that coheres with that semantic interpretation. There is also something it *feels like* to mean “the cat is on the mat.” And without feeling something like that, all there is is doing.

Now explaining how and why we feel rather than just do is the hard problem, whether it pertains to sensing

objects or meaning/ understanding sentences. I'd call that a profound explanatory gap, rather than an "epiphenomenon."

(But perhaps all that Professor Campbell meant by "epiphenomenon" here was that what it feels like to be saying and meaning a sentence is [roughly] what it feels like to be imaging or otherwise "calling to mind" its referents. I'd call that feeling "derivative" rather than "epiphenomenal," but that's just a terminological quibble, as long as we agree that meaning must be not only grounded but felt.)

John Campbell

[April 22nd, 2011 at 11:09 am](#)

Stevan speaks of ' the "symbol grounding problem" (one of the "easy" problems) The solution to the symbol grounding problem is to ground the internal symbols (words) of a Turing robot in its autonomous sensorimotor capacity to detect, categorize, manipulate and describe the symbols' external referents. But although grounding is necessary for meaning, it is not sufficient. The other necessary component is feeling'.

You can't address the symbol-grounding problem without looking at relations to *sensory awareness*. Someone who uses, e.g., words for shapes and colors, but has never had experience of shapes or colors, doesn't know what they're talking about; it's just empty talk (even if they have perceptual systems remote from consciousness that allow them to use the words differentially in response to the presence of particular shapes of colors around them). Symbol-grounding shouldn't be discussed independently of phenomena of consciousness.

Trying to leave out problems of consciousness in connection with symbol-grounding, and then bring them back in with the talk of 'feeling', makes for bafflement. If you stick a pin in me and I say 'That hurt' is the pain itself the feeling of meaning? The talk about 'feeling of meaning' here isn't particularly colloquial, but it hasn't been given a plain theoretical role either.

Stevan Harnad

[April 22nd, 2011 at 5:23 pm](#)

UNFELT GROUNDING: Reply to John Campbell-2

JC: "You can't address the symbol-grounding problem without looking at relations to sensory awareness. Someone who uses, e.g., words for shapes and colors, but has never had experience of shapes or colors, doesn't know what they're talking about; it's just empty talk (even if they have perceptual systems remote from consciousness that allow them to use the words differentially in response to the presence of particular shapes of colors around them). Symbol-grounding shouldn't be discussed independently of phenomena of consciousness."

The symbol grounding problem first reared its head in the context of John Searle's Chinese Room Argument. Searle showed that computation (formal symbol manipulation) alone is not enough to generate meaning, even at Turing-Test scale. He was saying things coherently in Chinese, but he did not understand, hence mean, anything he was saying. And the incontrovertible way he discerned that he was not understanding was not by noting that his words were not grounded in their referents, but by noting that he had no idea what he was saying — or even that he was saying anything. And he was able to make that judgment because he knew what it felt like to understand (or not understand) what he was saying.

The natural solution was to scale up the Turing Test from verbal performance capacity alone to full robotic performance capacity. That would ground symbol use in the capacity for interacting with the things the symbols are about, Turing-indistinguishably from a real human being, for a lifetime. But it's not clear whether that would give the words meaning, rather than just grounding.

Now you may doubt that there could be a successful Turing robot at all (but then I think you would have to explain why you think not). Or, like me, you may doubt that there could be a successful Turing robot unless it really did feel (but then I think you would have to explain — as I cannot — why you think it would need to feel).

If I may transcribe the above paragraph with some simplifications, I think I can bring out the fact that an explanation is still called for. But it must be noted that I am — and have been all along — using "feeling" synonymously with, and in place of "consciousness":

***IC: You can't address the symbol-grounding problem without looking at relations to feeling. A Turing robot that uses*

words for shapes and colors, but has never felt what it feels like to see shapes or colors, doesn't know what it's talking about; it's just empty talk (even if it has unfelt sensorimotor and internal systems that allow it to speak and act indistinguishably from us). Symbol-grounding shouldn't be discussed independently of feeling."

I think you are simply assuming that feeling (consciousness) is a prerequisite for being able to do what we can do, whereas explaining how and why that is true is precisely the burden of the hard problem.
You go on to write the following (but I will consistently use "feeling" for "consciousness" to make it clearer):

JC: *"Trying to leave out problems of [feeling] in connection with symbol-grounding, and then [to] bring [it] back in with the talk of 'feeling', makes for bafflement. If you stick a pin in me and I say 'That hurt' is the pain itself the feeling of meaning? The talk about 'feeling of meaning' here isn't particularly colloquial, but it hasn't been given a plain theoretical role either."*

I leave feeling out of symbol grounding because I don't think they are necessarily the same thing. (I doubt that there could be a grounded Turing robot that does not feel, but I cannot explain how or why.)

It feels like one thing to be hurt, and it feels like another thing to say and mean "That hurt." The latter may draw on the former to some extent, but (1) being hurt and (2) saying and meaning "That hurt" are different, and feel different. The only point is that (2) feels like something too: that's what makes it *meant* rather than just grounded.

Harnad, S. (1990) [The Symbol Grounding Problem](#) *Physica D* 42: 335-346.

John Campbell

[April 23rd, 2011 at 11:46 am](#)

'I think you are simply assuming that feeling (consciousness) is a prerequisite for being able to do what we can do, whereas explaining how and why that is true is precisely the burden of the hard problem.'

I think there are two different problems here:

- (1) Characterizing the epistemic role of consciousness. In particular, there's explaining the work that sensory experience does in (a) our having propositional knowledge of our surroundings, knowing that things are thus-and-so around us, and (b) having concepts of the objects and properties in our surroundings, knowing which objects and properties those are;
- (2) Explaining how conscious experience can be realized by a physical system.

It seems to me that (1) is not well understood, and that arguably it's prior to (2). I don't think there's much hope for a successful assault on (2) unless we have firmly in place a clear conception of exactly what explanatory work the notion of consciousness in general, and of sensory experience in particular, is doing for us.

[Stevan Harnad](#)

[April 23rd, 2011 at 7:18 pm](#)

THE EVER-ELUSIVE CAUSAL STATUS OF FEELING (Reply to John Campbell-3)

JC: *"(1) Characterizing the epistemic role of consciousness. In particular, there's explaining the work that sensory experience does in (a) our having propositional knowledge of our surroundings, knowing that things are thus-and-so around us, and (b) having concepts of the objects and properties in our surroundings, knowing which objects and properties those are"*

The trouble is that each of the mental states you mention has an easy aspect (doing and ability to do) and a hard aspect (feeling). So unless you specify which of the two you are referring to, it is difficult to know what you really mean:

JC: *"(1) Characterizing the epistemic role of consciousness"*

"Epistemic" is equivocal: it could refer to what can be known in the sense of unfelt knowing (doing, and ability to do: easy) or felt knowing (hard).

And until/ unless there are further arguments to show that the distinction is coherent, a “conscious” state is a state that it feels like something to be in, hence a felt state.

JC: *“In particular, there’s explaining the work that sensory experience does”*

Unfelt sensory system activity (doing, and ability to do: easy) or felt sensory experience (hard)?

JC: *“in (a) our having propositional knowledge of our surroundings”*

Unfelt propositional knowledge (doing and saying, and ability to do and say: easy) or felt knowledge (hard)?

JC: *“knowing that things are thus-and-so around us”*

Unfelt knowing (doing, and ability to do: easy) or felt knowing (hard)?

JC: *“and (b) having concepts of the objects and properties in our surroundings”*

I have no idea what “having concepts” means! Does it mean being able to do/say certain things (easy) or does it also feel like something to have a concept (hard)?

JC: *“knowing which objects and properties those are”*

Unfelt knowing (doing, and ability to do: easy) or felt knowing (hard)?

JC: *“(2) Explaining how conscious experience can be realized by a physical system. It seems to me that (1) is not well understood, and that arguably it’s prior to (2)”*

I agree.

JC: *“I don’t think there’s much hope for a successful assault on (2) unless we have firmly in place a clear conception of exactly what explanatory work the notion of consciousness in general, and of sensory experience in particular, is doing for us”*

I agree. And the hard part is that on the face of it the answer is: none!

Galen Strawson

April 25th, 2011 at 8:29 am

Dear Stevan, I find it hard to accept the use of the word ‘feel’ to cover the case of cognitive phenomenology (cognitive experience, ‘meaning-experience’, ‘understanding-experience’, ‘semantic phenomenology’). Apart from that I think I agree with everything you say.

Maybe we can reach an accommodation. In a 1994 book I say “my central claim is that the apprehension and understanding of cognitive content, considered just as such and independently of any accompaniments in any of the sensory-modality-based modes of imagination or mental representation, is part of experience, part of the flesh or content of experience, and hence, trivially, part of the qualitative character of experience.” (Mental Reality p. 12; new edition 2009) It seems to me that when you say that

“It feels like something to mean something. It also feels like something to think, believe, understand or doubt something”

you may mean no more than that it’s “part of the qualitative character of experience”. (I understand the word ‘experience’ in such a way that it’s true by definition that all experience has experiential qualitative character.)

You may alienate people who would otherwise agree with you if you speak wholly in terms of ‘feeling’. I think we do best to establish the maximally general category of an ‘experiential modality’. We can then bring the familiar notion of a sensory modality or sense/feeling modality under it, and leave open the possibility that there may be experiential modalities whose activity is of course a matter of experience, i.e. of experiential qualitative character, but

is nevertheless not best thought of as a matter of feeling. If you identify the notion of experiential qualitative character with that of feeling, then we agree on the facts, and disagree only on the terminology.

I think the terminological situation can be summarized as follows: Some take it to be true by definition that [a] all feeling is sensory feeling (where 'sensory' includes mood feelings) and of course conversely

These people may well agree with you that

[b] that all experience properly speaking is just a matter of feeling

but this will be because they hold (falsely in my view) that

[c] all experience properly speaking is just a matter of sensory feeling and deduce [b] from [c] and [a].

I think [a] is a reasonable way to understand feeling, but I reject [c], as I think you do, because I accept [a], and so reject [b]. You, I take it, reject [a] and [c], and accept [b]. I think you accept [b] because you take it to be true by definition that

[d] all experiential qualitative character is a matter of feeling

from which [b] follows, given the trivial (definitional) truth that

[e] all experience is a matter of experiential qualitative character.

I reject [d], so although I accept [e], I don't infer [b]. But if you tell me that [d] is true by definition for you, then I agree with your substantive position. I just don't think it's the best way to put things.

None of this is any help with the 'hard problem', but the hard problem rests essentially on a false assumption: the assumption that we know something about the nature of the physical that gives us a good reason to think that there is a problem in the idea that the experiential is physical. This assumption is false.

I don't know if this is of any help. In conclusion, another passage from the 1994 book (p. 196): "Each sensory modality is an experiential modality, and thought experience (in which understanding-experience may be included) is an experiential modality to be reckoned alongside other experiential modalities. We have, so far, no explanation of how the systems of the eye and brain give rise to the phenomenology of color experience in the particular way that they do (leaving aside partial "abstract morphology" explanations like the Churchland-type explanation mentioned in section 4.2). In the same way, we have no explanation of how the systems of the brain that underlie or realize thought give rise to, or involve, conscious thought experience in the way in which they do. The fact remains that our cognitive lives are, as such, experientially rich. This is perhaps never more apparent than when one is lying in the dark thinking of one thing after another, unable to sleep. (Insomnia has philosophical uses.)"

Stevan Harnad

April 25th, 2011 at 11:45 am

DIVERGING ON TERMS, CONVERGING ON SUBSTANCE (Reply to Galen Strawson)

GS: "If you identify the notion of experiential qualitative character with that of feeling, then we agree on the facts, and disagree only on the terminology."

Then we agree on the facts and just disagree on the terminology!

(I find it much more straightforward and natural to speak about what experiences feel like than to speak of their "qualitative character" — but absolutely nothing substantive rides on this taste in terms.)

GS: "the hard problem rests essentially on a false assumption...that we know something about the nature of the physical that gives us a good reason to think that there is a problem in the idea that the experiential is physical"

My hard problem is not that metaphysical one, but this epistemic one: Ee cannot explain how and why we feel rather than just do (or, if you wish, why and how we have "experiences with qualitative character" rather than just do).

If I may translate into my preferred terms the paragraph you quote from Strawson (1994) (p. 196):

Each sensory experience is felt, and each thought experience is felt. We have, so far, no explanation of how the eye and brain give rise to feeling. In the same way, we have no explanation of how the systems of the brain that generate thought give rise to feeling. The fact remains that we feel.

I agree that we have no explanation “so far.”

(I also give some reasons in my paper why I don't think we ever will. Among other things, I think your own preferred “panpsychism” pays far, far too exorbitant an ontic price for very little in the way of an explanatory purchase. It hypothesizes, without evidence, that feeling is a ubiquitous latent feature of matter all over the universe — which, amongst other things, creates a bit of a mereological nightmare — leaving it just as much of a mystery how and why we feel rather than just do. It borrows the bottom-line — the-buck-stops-here — character of the fundamental forces [electromagnetic, gravitation, strong subatomic, weak subatomic], but without their massive supporting evidence or explanatory power.)

Bernard Baars

April 25th, 2011 at 2:59 pm

THE BRAIN IS NOT A TURING MACHINE!

I realize that traditionally Turing Machines are taken to be abstract versions of all possible computational implementations, including bio computation. If you can therefore prove, or quasi-prove, that something is possible or impossible for a Turing Machine that is taken to apply to all possible computers.

The trouble is that the assumption is wrong.

1. Turing Machines have no memory, and no time, and no string limits. Those are non-biological assumptions.
2. Turing Machines are rigidly serial, when the brain is a massively parallel, and parallel-interactive organ.
3. While it is argued that TM's can simulate parallel and parallel-interactive computations, that is plausible only because TM's totally ignore memory, time, and finite string limits.
4. I believe that Stan Franklin and a colleague have given a formal proof that contrary to earlier claims, there are formal machines that are more powerful mathematically than Turing Machines. This vitiates the whole standard use of TMs.
5. Consciousness and qualia are biological entities, which are selectionist rather than instructionist in principle (GM Edelman), and reflect a huge evolutionary history — 200 million for mammals alone.
6. We have a long and repeated history of ‘impossibility proofs’ designed to falsify important empirical advances. Newton's action at a distance, the molecular basis of life, etc. These efforts routinely fail, though they sometimes do so in interesting ways.
7. There is no substitute for looking at nature.

Bernard Baars

Stevan Harnad

April 25th, 2011 at 7:11 pm

A TURING ROBOT IS NOT A TURING MACHINE (Reply to Bernard Baars)

I don't think anyone on any side of this discussion has said that the brain is a Turing Machine. The one who comes closest, Shimon Edelman, explicitly says “I argue that feelings in fact are computations, albeit not Turing computations.”

A Turing robot (i.e., a robot capable of passing the Turing Test, indistinguishably from any of the rest of us, for a lifetime) is not a computer (Turing machine). It is a dynamical system, with sensors and effectors, and on the inside it may be implementing any processes — whether dynamic or computational — that give it the capacity to pass the Turing Test, Turing computation being only one among the many possible processes.

The “weak” version of the Church-Turing Thesis is that everything that is “effectively computable” for a mathematician is computable by a Turing Machine.

The strong version of the Church-Turing Thesis is that Turing computation (digital computation) can simulate and approximate (just about) any dynamical physical process in the universe, including sensors and effectors, as well as analog continuous, parallel, distributed processes (such as internal rotation), and indeed also just about any neuro-chemical brain processes (perhaps excluding quantum and chaotic processes). But that simulation is only formal. A purely computational airplane does not fly. And a purely computational brain does not cognize (nor, a fortiori, does it feel). Nor does a purely computational robot (a “virtual robot”).

It is an empirical question, however, what and how much of the actual internal functioning of a Turing robot (or brain) could be performed by Turing computation.

What's sure is that it cannot be all of it.

BB: *“I realize that traditionally Turing Machines are taken to be abstract versions of all possible computational implementations, including bio computation. If you can therefore prove, or quasi-prove, that something is possible or impossible for a Turing Machine that is taken to apply to all possible computers. The trouble is that the assumption is wrong.”*

The strong version of the Church-Turing Thesis holds that Turing computation can simulate and approximate (just about) any dynamical physical process — not that it can stand in for any dynamical physical process. You can't fly to Chicago on a simulated airplane; flying is not computation. But computation can decompose and test the causal explanation of flying (or cognition).

BB: *“1. Turing Machines have no memory, and no time, and no string limits. Those are non-biological assumptions.”*

Turing machines are formal abstractions, but they can be implemented in real finite-state dynamical systems, for example, digital computers (which do have memories, clocks and length limits).

BB: *“2. Turing Machines are rigidly serial, when the brain is a massively parallel, and parallel-interactive organ.”*

Yes, but as noted, nobody says the brain is a Turing machine, just that the brain can be simulated computationally by a Turing machine.

BB: *“3. While it is argued that TM's can simulate parallel and parallel-interactive computations, that is plausible only because TM's totally ignore memory, time, and finite string limits.”*

They can simulate them because the parallelism is simulated serially, in virtual rather than real time.

BB: *“4. I believe that Stan Franklin and a colleague have given a formal proof that contrary to earlier claims, there are formal machines that are more powerful mathematically than Turing Machines. This vitiates the whole standard use of TMs.”*

The subject of hypercomputation is controversial and I think the “hard” problem of explaining feeling is hard enough without complicating it with speculations about hypercomputation (or quantum mechanics!). The weak Church-Turing Thesis stands unrefuted to date: Whatever mathematicians have regarded as computation has turned out to be Turing machine-computable. The strong Church-Turing Thesis does not hold that everything is computer-simulable, only just-about everything.

BB: *“5. Consciousness and qualia are biological entities, which are selectionist rather than instructionist in principle (GM Edelman), and reflect a huge evolutionary history — 200 million for mammals alone.”*

No doubt. But feeling (i.e., consciousness, qualia) poses a special, hard *hard* problem, both for evolutionary explanation and for functional/causal explanation. This problem will be the subject of the 2012 Summer School on the Evolution and Function of Consciousness at the Université du Québec à Montreal in June/July 2012 in which many of the contributors to this discussion (including Bernie Baars) and many other thinkers will be participating. (The Summer School will also be in commemoration of the centennial of Turing's birth in June 1912).

BB: *“6. We have a long and repeated history of ‘impossibility proofs’ designed to falsify important empirical advances. Newton's action at a distance, the molecular basis of life, etc. These efforts routinely fail, though they sometimes do so in interesting ways.”*

Explaining how and why we feel is hard (indeed, I think, impossible), but the reason has nothing to do with Turing machines or computation, nor with either the weak or the strong Church-Turing Thesis. (See “Vitalism, Animism and Feeling (Reply to Anil Seth)” in this discussion.)

BB: *“7. The brain is a complex system of interacting parts, and the whole is more than the sum of its parts.”*

BB: "... There is no substitute for looking at nature."

Logic is an ineluctable part of nature too...

Harnad, S. (2008) [The Annotation Game: On Turing \(1950\) on Computing, Machinery and Intelligence](#). In: Epstein, Robert & Peters, Grace (Eds.) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer

Krisztian Gabris

[April 25th, 2011 at 5:41 pm](#)

Stevan's question was to bring up facts or speculative evidence that could possibly explain how and why we feel rather than just do.

That is, how and why the brain generates feelings (subjective experiences, or qualia) instead of just being an exact functional equivalent with "no one home". Running a motor program of pulling away an arm to minimize tissue damage, without actually feeling anything in the meantime (for example).

It is rather speculative, but one approach is to try to think what could be the evolutionary adaptive advantage of such mental states.

Take the pain example. What would be the motivation to pull away the arm? It would be a long evolutionary selection process, that makes sure that when there is a pain signal, the arm is taken away, so there is no damage, or even death. But what would happen if for some reason in a decision point a decision is made which goes against the evolutionary ingrained rules of the system. For example, a hand is left in the fire. Let's suppose, that such behavior could emerge randomly in complex systems like a Turing robot, and it is not inherent of a certain genetic configuration (it cannot be selected out). What would be the punishment of such behavior in a Turing robot (other than tissue damage)? Nothing, the robot would go on it's own business with signals and internal warnings, but it would not feel the pain. Whereas a human would subjectively feel pain, and would take away the hand (except in cases of proving their trust in somebody) not only because of programming, but because of more immediate reasons of feeling pain.

It is rather speculative, but the main point I try to make is whether it is possible to think of a motivational value of feelings which could be adaptive (to motivate the following of the evolutionary selected behavior-program).

The weak point of the speculation is the randomly emerging behavior, which goes against evolutionary programming. What I had in mind is that perhaps complex systems are more error prone than we might think, and the evolutionary rules need support from subjective experiences (of pain for instance) to make sure an adaptive behavior is followed.

[Stevan Harnad](#)

[April 26th, 2011 at 11:18 am](#)

WHY A DISPOSITION TO FEEL AND THEN TO DO AS YOU FEEL — RATHER THAN JUST A DIRECT DISPOSITION TO DO? (Reply to Krisztian Gabris)

KG: "Take the pain example... what would happen if for some reason... a decision is made which goes against the evolutionarily ingrained rules of the system. For example, a hand is left in the fire... What would be the punishment of such behavior in a Turing robot (other than tissue damage)? Nothing, the robot would go on it's own business with signals and internal warnings, but it would not feel the pain. Whereas a human would... feel pain, and would take away the hand... not only because of [genetic] programming, but because of... feeling pain."

Yours is the natural intuitive explanation for why we feel — the one that feels right. "Why," after all, is a causal question: Why do we pull our hand out of the fire? Yes, fire causes tissue damage, but that's not what makes us withdraw our hand (unless we are anaesthetized): It's because it hurts!

So surely that's what pain's for: To signal tissue damage by causing pain to be felt.

Why? So you'll withdraw your hand. Because if your ancestors had been indifferent to tissue damage, they would not have had surviving descendants.

So you withdraw your hand because it hurts. And it hurts in order to cause you to feel like withdrawing your hand — and therefore you withdraw your hand.

Iniurv -> pain -> withdraw hand

And the reason the feeling of pain evolved is because those whose ancestors felt pain were more likely to feel like withdrawing their hands than those who did not.

But let us note that what was needed, for survival, was to withdraw the injured hand — an act, not a sentiment. The pain was a means, not an end. It's an extra step; and, as I will try to illustrate with other examples, a superfluous extra step, practically speaking. So the hard problem is to explain how and why this extra, apparently superfluous step evolved at all.

Suppose that what you had chosen for your evolutionary example of the adaptive trait for “motivational” scrutiny had been — rather than the withdrawing of the injured hand — the growing of wings, or the beating of the heart or the dilating of the pupil of the eye.

You'll perhaps find it strange to ask about feeling the “motivation” to grow wings (though it's a reasonable question), because growing is not something we ordinarily think of ourselves as “doing.” But note that the very same question you asked about the evolution of pain — and the “punishment” for non-withdrawal of the injured hand if no one feels the “motivation” to withdraw it — applies to the non-growth of wings. And the answer is the same:

If we are talking about evolution — which means traits that increase the likelihood of survival and reproduction — then for both the disposition to grow wings and the disposition to withdraw the hand from injury the “reward” is increased likelihood of survival and reproduction; and for both the lack of the disposition to grow wings and the lack of the disposition to withdraw the hand from injury the “punishment” is decreased likelihood of survival and reproduction.

The very same evolutionary reward/punishment scenario also applies to the disposition of our hearts to beat which is even more obviously something that our bodies *do* — or, if you want an example of something we do in response to a circumstantial stimulus rather than constantly, there's pupillary dilation to light intensity. Or, if you want something we do voluntarily rather than involuntarily — although that's begging the question, because it is really the involuntary/voluntary distinction that poses the “hard” problem and calls for explanation — consider the implicit improvement in skills that occurs without any sense of having done anything deliberately (sometimes even without the feeling that we have improved) in implicit learning, or the changes in our dispositions caused by subtle Pavlovian conditioning or Skinnerian reinforcement when we don't even feel that our dispositions are changing, or the voluntary take-over of breathing — usually involuntary, like the heart-beat.

And a disposition is a disposition to do, whether it's to grow, to beat, to dilate to withdraw, to salivate, to smile or to breathe. So the question remains: Why the extra intermediate step of feeling, when the reward and punishment come from the disposition to do?

The very same reasoning applies to learning itself: We learn to do things — such as what to eat and what to avoid — by trial and error and reward/punishment. The consequences of doing the right thing feel good and the consequences of doing the wrong thing feel bad, so we learn to do the right thing. “Motivation” again. But again, it is the *disposition to do* the right thing that matters; the feeling of reward and punishment is an extra. Why? Both in evolution and in learning there are consequences (enhanced survival and reproduction in the case of evolution, and enhanced functioning and performance in the case of learning: eating nourishing things gives us energy, eating toxic things makes us sick) and the consequences are sufficient to guide our dispositions to do. But why is any of that felt rather than just done?

These questions are hard not only because of the underlying problem of causality, but because our intuitions keep telling us that it's obvious that we need to feel. Yet the causal role of feeling is anything but obvious, if looked at objectively, which means functionally.

You assumed that a Turing robot would not feel. That's not at all sure. But let's consider today's rudimentary robots, which are as unlikely to feel as a toaster or a stone. Yet even they can already be designed to withdraw damaged limbs, or to learn to withdraw damaged limbs. They need sensors, of course, but it's not at all clear why they would need feelings (even if we had the slightest clue of how to design feelings!), if the objective is to do — or to learn to do — what needs to be done in order to survive and function. They need to *detect* tissue damage, and then they need to be disposed to do — or disposed to learn to do — whatever needs to be done. If (sensible) anti-Creationism impels us to reject arguments from robotic design, consider that in evolution can be simulated computationally in artificial life simulations; and the kinds of traits we build into our robots can therein be shown to evolve by random variation and selection; the same can be done for computer models of learning (which just involve a change in simulation time scale), including computer models of the evolution of the disposition to learn (e.g., Baldwinian evolution).

And lest we propose the superior power of cognition over Pavlovian and Skinnerian learning, remember that the kind of information processing underlying cognition can be implemented (along with its power and benefits) computationally, in unfeeling machines.

So there is definitely a problem here, of explaining the ostensibly superfluous causal role of feeling in doing.

And not only do our intuitions fail us, but so does every objective attempt at the kind of causal explanation that serves us so well in just about every other functional dynamic under the sun.

To be continued in the [2012 Summer School on the Evolution and Function of Consciousness...](#)

Shikha Singh

[April 26th, 2011 at 8:40 pm](#)

In regards to meaning, feeling, and explaining it is quite difficult to figure out how or why we do it. What comes into my mind is the idea of intuition or 'gut feeling.' When someone feels that it is a bad idea to get on a certain plane or it is a bad idea to go out we blame it on 'gut feeling' or intuition. But how or why exactly do humans get such feelings? Where does it come from?

As mentioned in the essay it is simply easy to explain why or how we do things the way we do by creating the biorobot. However, could we not extend this biorobot to feeling? Why is it we cannot say the robot also has feelings? Is it because the robot does not have a human brain or any brain that animals or human beings do? Could we not say that the "motherboard" of the robot or the controller is the brain of the robot? Can we not say that we ourselves are robots in a way? In the essay above, feeling, meaning, and explaining are all in a sense combined to meaning according to the philosophical perspective. If this is the case, then can one not say that feeling is all else that is not doing? Hence, the word feeling is used to describe all other actions that cannot be designated to the doing category. Therefore, in a sense, there are now only two things one can do-feel or do. We designate all our other actions of explaining, meaning, in tuition, 'gut feeling,' and all other to the word feeling. Now the question that comes to mind is if feeling is everything but doing then how or why is it that we feel? Imagine that the biorobot is doing something and is notified that its power is running low for some reason so it decides to shut down. Why did it decide to shut down? Can we acknowledge this as feeling? Why not? One can easily argue that the biorobot was programmed to shut down and therefore it has nothing to do with feeling, but then are us humans also not just programmed to 'feel' certain ways in certain scenarios by the rules of societies?

Take for example Genie the "fear child," and consider her behavior. Genie was locked in and put far away from society for 12 years. The only human contact she had was from her parents. With such limitation, Genie's personality was different and she was completely mute at one point as if she had no desires, no feelings. However, once opened up to the society and shown the outside world, Genie began showing signs in relevance to normal human behavior. Why is that? Is it because it was within her all along or did someone had to guide her or "program" her? This is something to think about and it could be noted that feeling is merely another part of society's requirements and thus that is why we do it and although humans are not willing to say they are robots, in some way we can be seen as robots that are programed which is how we feel.

Stevan Harnad

[April 26th, 2011 at 10:13 pm](#)

HOME TRUTHS ABOUT FEELING, DOING, EXPLAINING AND ROBOTS (Reply to Shikha Singh)

Doings are observable by anyone (via senses or senses plus measuring instruments).

Feelings are observable only to their feeler.

The only feelings a feeler can feel are his own.

That other people and animals feel is a safe guess, because they are related to and resemble us.

That today's man-made robots feel is as unlikely as that a toaster or stone feels.

That a robot whose doings are Turing indistinguishable from the rest of us for a lifetime would feel would be almost as safe a guess as that other people and animals feel. (Perhaps a biorobot would be an even safer guess).

A robot is just an autonomous causal system that can do some things that people and animals can do.

Cognitive science is about discovering the causal mechanism that generates our capacity to do what we can do.

(We can think of it as discovering what kind of robots we are.)

No one but the Turing robot can know whether its causal mechanism does generate feeling.

And even if it does, not even the Turing robot can explain or know how or why.

Galen Strawson

April 27th, 2011 at 9:25 am

SH: I think your own preferred “panpsychism” pays far, far too exorbitant an ontic price for very little in the way of an explanatory purchase. It hypothesizes, without evidence, that feeling is a ubiquitous latent feature of matter all over the universe

GS: In order to reply, I'll assume the truth of materialism/ monism:

[1] materialism/ monism is true.

And I'll assume the real unproblematic reality of consciousness/ experience/ experientiality/ feeling

[2] consciousness/ experience/ experientiality/ feeling is real.

The first thing to note, perhaps, is that

[3] *there is absolutely no evidence that anything non-experiential exists* (nor will there ever be).

So there is to that extent no evidence that panpsychism is ontically costly. It's true that the explanatory power of physics (the human creation) doesn't explicitly appeal to the idea of experientiality, but if in fact energy is experientiality—if in fact the intrinsic nature of what we detect as energy is experientiality—if, for short,

[4] everything physical is experiential (= panpsychism is true)

then physics is in fact talking about experientiality all along. And, crucially,

[5] the explanatory power of physics *doesn't in any way conflict* with [4],

and [4] solves the 'hard problem' at a high level of generality—albeit without any prospect of their being any detailed useful explanation of the phenomenon of experientiality. So why not be open to the hypothesis that [4] is true?

Note that few people think we need an explanation of the existence of *non*-experientiality (ie of the physical as ordinarily conceived). Why should one think that one needs an explanation of the existence of experientiality? It is widely agreed that physics deals only in equations and numbers and structures, and says absolutely nothing whatever about the intrinsic nature of that which has the structure (the physical), so far as that intrinsic nature is something more than its structure. Russell was very clear about this.

There is I think very great danger of *misusing the idea of explanatory power* (there is a mindset that makes it extremely hard to really think through the case for panpsychism. It can take years to break through.) But consider: assuming [1] materialism,

[6] we know that some parts of the physical are experiential

and, as [3] states, we don't know that any parts of it are non-experiential, and will never have any positive evidence that they are. One way of applying Occam's Razor leads straight from [6] to the view that [4] everything physical is experiential. For [6], we know that some parts of the physical are experiential, and we have, yet again, [3], no evidence that any part of the physical is non-experiential. So why should we positively suppose, in the absence of any supporting evidence, that the physical is or must be fundamentally different in different parts—i.e. deny [4], and hold that the physical is definitely partly non-experiential? Especially when [4] solves the hard problem at a high level of generality?

One argument is this. If

[7] 'radical emergence' is impossible

(there are potent reasons for thinking this is so) then, given [4], the fact that we know that at least some parts of the physical are experiential, we can it seems deduce that

[8] experientiality must be a fundamental property of the physical.

So that either panpsychism is true (experientiality is fundamental property of all parts of the physical), or

micropsychism is true (experientiality is fundamental property of some fundamental parts of the physical). If

[9] 'smallism' is true (the physical universe really does come in small bits)

and if

[10] there is only one truly fundamental kind of fundamental physical entity,

as many suppose, then it looks as if panpsychism wins out over micropsychism.

There is still, certainly, the mereological nightmare. Here I recommend William James's A Pluralistic Universe.

Stevan Harnad

April 27th, 2011 at 3:26 pm

A FIFTH FORCE: BUT AN ACAUSAL ONE... (Reply to Galen Strawson-2)

Galen Strawson does a brilliant, heroic job with panpsychism:

The only thing we know for sure — indeed, with a Cartesian certainty that is as apodictic as the logical necessity of mathematics — is *that* and *what* we feel.

Everything else we know (or believe we know), we likewise know “through” feeling — in that it feels like something to learn it and it feels like something to know it.

(It feels like something to make an “empirical” observation. It feels like something to understand that something is the case. It feels like something to understand an inference or a causal explanation.)

So feeling is certain, whereas physics (“doing,” in my parlance) is not certain.

But we are realists, trying to do the best we can to explain reality — not extreme sceptics, doubting everything that is not absolutely certain, even if it’s highly probable.

We are just looking for truth, not necessarily certainty.

“Experience” is a weasel-word because it can mean either feeling something — which is highly problematic (the “hard problem”) — or it can just mean acquiring empirical data (as in: “this machine had the solution built in, that machine learned it from experience”) — which is unproblematic (doing, the “easy” problem).

So whereas it is true that the only thing we know for sure (besides the things that are necessarily true on pain of contradiction) is that feeling exists, neither everyday life nor science requires certainty. High probability on the evidence (data) will do.

And although it is true that all evidence is felt evidence, it is only the fact that it is felt that is certain. The evidence itself (doing) is only probable.

In other words, although they always accompany the data-acquisition (doing), the feelings are fallible. We feel things that are both true and untrue about the world, and the only way to test them out is via doings. It is true that the data from those doings are also felt. But the felt data are answerable to the doings, and not to the fact that they are felt.

And not only are our feelings fallible, as regards the truth: they also seem to be causally superfluous. Doings (including data-acquisition) alone are enough, for evolution, as well as for learning. Some doings are undeniably felt, but the question is: how and why?

When we are doing physics (or chemistry, or biology, or engineering) and causal explanation (rather than metaphysics), we have to explain the facts, amongst which one fact — the fact that we feel — seems pretty refractory to any sort of explanation except if we suppose that feeling is simply a basic property of the universe (whether local to the organisms in the earth’s biosphere [Galen’s “micropsychism”] or somehow smeared all over the universe [“panpsychism”).

There’s no doubt that feeling exists, so in that sense feeling is indeed a property of the universe. But with all other properties — doings, all — we have become accustomed to being able (in practice, or at least in principle) to give a causal explanation of them in terms of the four fundamental forces (electromagnetism, gravitation, strong subatomic, weak subatomic). Those forces themselves we accept as given: properties of the universe such as it is, for which no further explanation is possible.

Galen’s metaphysics would require adding something like a fifth member to this fundamental quartet — feeling — with the difference that, unlike the others, it is not an independent force, it does not itself cause and thereby explain doings causally, but rather is merely correlated with them, inexplicably, for some doings.

And our justification for adding a fifth acausal force? The fact that it is inexplicably (but truly) correlated with some doings (all doings that we feel). If feeling had truly been a 5th force (causal rather than acausal), namely, “psychokinesis” (“mind over matter”), then that would indeed have merited elevating it to fundamental status, exempt from further explanation along with the other four.

But there is not a shred of evidence for psychokinesis as a causal force (and all attempts to measure psychokinesis have failed, because the other four forces already covered all the causal territory — doing — with no remainder and no further room for causal intervention).

So all we have, inexplicably, is the fact that we feel. I don’t think that that fact warrants any further metaphysics than that: feeling definitely exists — and, unlike anything else, exists with certainty rather than just probably. It also happens to feel like something to find out and understand anything we know. The rest is an epistemic problem: why and how does getting or having data feel like something (for feeling creatures like us)?

Neither “micropsychism” nor “panpsychism” answer this question. They just take it for granted that it is so.

April 29th, 2011 at 3:01 pm

TALKING ABOUT FEELING: SUMMARY OF FORUM

In my little [essay](#) I tried to redraft the problem of consciousness — the “mind/body problem” — as the problem of explaining how and why we feel rather than just do.

It was not meant as a terminological exercise. The usual way we talk about consciousness and mental states uses weasel-words (“conscious,” “mental,” “experience”) that are systematically ambiguous about whether we are just talking about access to data (an easy problem, already solved in principle by computation, which is simply an instance of doing) or about *felt* access to data (the hard part being to explain not just the doing but the feeling). Nor was it meant as a metaphysical exercise: The problem is not one of “existence” (feeling indubitably exists) but of *explanation*: How? Why?

The commentaries were a fair sample, though a small one, of the issues and the kinds of views thinkers have on them today. A much fuller inventory will be presented at the [2012 Summer School on the Evolution and Function of Consciousness](#) in Montreal June/July of next year. Think of this small series of exchanges in the *On the Human* Forum as an overture to that fuller opus.

I have already responded in detail individually to each of the 10 commentators (15 commentaries) so I will just summarize the gist here:

Judith Economos rightly insists, as the only one with privileged access to what’s going on in her mind, that it is not true that she feels everything of which she is conscious: Some of it — the part that is not sensory or emotional — she simply *knows*, though it doesn’t feel like anything to know it. I reply (predictably) that “know,” too, is a weasel-word, ambiguous as between felt and unfelt access to data. So if one is awake (conscious) whilst one is knowing, one is presumably feeling *something*. One is also, presumably, feeling something whilst one is not-knowing something, or knowing something else. If all three of those states feel identical, how does one know the difference? For if “knowing” just refers to having data, then it is just a matter of **know-how** (doing), which is already explained (potentially) by computation, and has nothing to do with consciousness.

Galen Strawson seems to agree with me on the distinction, but prefers “experience” (“with qualitative character”) to “feeling.” Fine — but “experience” alone is ambiguous; and trailing the phrase “with qualitative character” after it seems a bit burdensome to convey what “feel” does in one natural, intuitive, monosyllabic swoop. The substantive disagreement with Galen is about the coherence and explanatory value of “panpsychism” (i.e., the metaphysical hypothesis that feeling, or the potential to feel, is a latent and ubiquitous property of the entire universe) as a solution to the hard problem. The existence of feeling is not in doubt. But calling it a fundamental take-it-or-leave-it basic property of the universe does not explain it; it’s just a metaphysical excuse for the absence of an explanation!

Shimon Edelman is more optimistic about an explanation because there are computational and dynamic ways to “mirror” every discriminable difference (JND) in a system’s input in differences in its internal representations. This would certainly account for every JND a system can discriminate; but discrimination is *doing*: The question of how and why the doing is felt is left untouched.

David Rosenthal interprets the experimental evidence for “unconscious perception” as evidence for “unconscious feeling,” but, to me, that would be the same thing as “unfelt feeling”, which makes no sense. So if it’s not feeling, what is unconscious “perception”? It is unconscious detection and discrimination — in other words, internal data-doings and dispositions that are unproblematic because they are unfelt (the easy problem). If all of our know-how were like that, we’d all be Zombies and there would be no hard problem. David needs unconscious perception to be able to move on to higher-order consciousness (but that is, of course, merely higher-order *access* — the easy part, until/unless feeling itself is first explained). So this seems like recourse to either a bootstrap or a skyhook.

John Campbell points out that sensorimotor grounding is not enough to explain meaning unless the sensing is felt, and I agree. But he does not explain how or why sensorimotor grounding is felt.

Anil Seth reminds us that many had thought that there was a “hard problem” with explaining life, too, and that that turned out to be wrong. So there’s no reason not to expect that feeling will eventually be explained too. The trouble is that apart from the observable properties of living things (“doings”) *there was never anything else that vitalists could ever point to*, to justify their hunch that life was inexplicable unless one posited an “*elan vital*.” Modern molecular biology has since shown that all the observable properties of life could be explained, without remainder, after all. But in the case of feeling there *is* a property to point to — observable only to the feeler, but as sure as anything can be — that the full explanation of the observable doings leaves out and hence cannot account for. (Perhaps feeling is the property that the vitalists had in mind all along.)

The remaining commentaries seem to be based on misunderstandings:

Bernard Baars took “Turing Robot” to refer to “Turing Machine.” It does not. A Turing Machine is just a

formalization of computation. The internal mechanism of a Turing Robot can be computational or dynamical (i.e., any physical process at all, including neurobiological).

Krisztian Gabris thinks feelings are needed to “motivate” us to do what needs to be done. That’s certainly what it feels like to us. But on the face of it, the only thing that’s needed is a *disposition* to do what needs to be done. That’s just know-how and doing, already evident in toy robots and toasters. How and why it (sometimes) *feels like something* to have a disposition to do something remains unexplained.

Joel Marks assumed that the Turing Robot would be an unfeeling Zombie. This is not necessarily true. (I think it would feel — it’s just that we won’t be able to know whether it feels; and even if it does feel, we will be unable to explain how or why.) Hence Joel’s question about whether it would be wrong to create a robot that feared death is equivocal: By definition, if it’s a Zombie, it cannot fear, it can only act as if it feared. (Witnessing that may make *us* feel bad, but the Zombie — if there can be Zombies — would feel nothing at all.) And if the Turing Robot feels, it’s as important to protect it from hurt as it is to protect any other feeling creature from hurt.

Jason King

[April 29th, 2011 at 3:12 pm](#)

This conversation, while ending here, continues on Facebook. Join us there by logging on to your Facebook account and proceeding to our group: [On the Human](#).

COMMENTARY BY ANTONIA CHELLA & RICCARDO MANZOTTI:

Is the ontological framework sound enough?

Antonio CHELLA,

DICGIM, University of Palermo
Viale delle Scienze building 6, 90128 Palermo (Italy)
chella ++ unipa.it

Riccardo MANZOTTI

Institute of Communication and Behavior “G. Fabris”
IULM University, 20143, Via Carlo Bo, 8, Milano, Italy
riccardo.manzotti ++ iulm.it

In his target paper, Stevan Harnad draws a series of inescapable conclusions based on the received view as to the contrast between doing and feeling. He is absolutely correct to the extent that such assumptions are correct. What is surprising is how little critical consideration is given to the robustness of such fundamental thesis. In this commentary, we’d like to wonder whether such a received framework is sound enough. In particular we are going to focus on the first pair of issues outlined by Harnad – namely doing vs. feeling.

It is the very strength of Harnad’s logic coupled with the empirical implausibility of his conclusions that suggests questioning his starting hypothesis. For instance, consider his statement that “the feeling is not explained by the causal mechanism of the doing”. Of course, it is true. However, why should we be surprised by it? Is doing all that matters? Or is there something else?

It seems that Harnad assumes that the “doing” exhaust the explanatory power of human understanding as if it were the only way we can explain and describe reality. This is a very functionalistic standpoint which maintains that a functional description is the only available one. However, since the fact that we feel is an empirical undeniable fact albeit from a first-person perspective, we should argue against any view that does not predict such possibility. If feeling would not fit into the functional description of reality, so much the worse for functionalism. Harnad seems to follow a totally different conceptual road. IN his opinion, a functional description of reality cannot be challenged – or so it seems to us. In fact he keeps contrasting explanations of how brains do things with how brains feel things as if only the former were an scientifically acceptable one.

It must be stressed that in the above paragraph we purposefully shifted from a causal description to a functional one since we suspect that Harnad drifted towards the latter. The reason is simple and we will outline our concerns by means of an example based on classic physics. Consider a mathematical/functional description for the behavior of two mass particles in the empty space. Their trajectories are definable in mathematical terms. Consider now the same kind of description for two electric particles with opposite charge. Functionally, the descriptions of the two cases are identical once the proper quantities are chosen. It is a well known fact that for limited speeds the equations for gravity and electromagnetism have the same form (a fact that puzzled and misled the late Albert Einstein). The two cases are functionally identical. Yet, they are different both in causal and in physical terms since the physical properties (or powers) which are responsible for the two situations are very different (on one hand, mass and gravity and, on the other hand, electric charge and electromagnetic force). This trivial example is sufficient to show that the functional description (what Harnad would dub the “doing”) is not enough to pinpoint the nature of physics. Something else is required.

Similarly, if two doings (in terms of trajectories of particles) have identical functional descriptions yet different physical nature, why should it be surprising that two doings (a behavior of a human being vs. a behavior in a Turing robot) may have different nature? Of course, so far, we are missing a complete description of reality as it is plain from the fact that feeling do not fit into the dominant scientific framework. Yet, is this a good reason to assume that it won’t ever be available? Or that the only acceptable strategy is a functionalistic one (which would fail for existing and relatively better understood phenomena such as electricity and gravity)?

It seems to us that Harnad’s approach assumes to know a lot more than it is actually known about the nature of the physical world. As Galen Strawson recently observed “We can know for sure that we’re quite hopelessly wrong about the nature of matter so long as our positive account of it creates any problem about how consciousness can be physical.” (Strawson 2011) What is still missing is a theory outlining a conceptual and

causal connection between neural activity and phenomenal experience and functionalism does not seem to possess the resources to do it. Thus it may be worth to consider other hypotheses however counterintuitive they may seem to be. After all, "If no theories seem to be capable of accounting for conscious experience, this probably means that there is something inherent in our assumptions that divorce theories from conscious experience" (Rockwell 2005, p. 49).

Another reason of concern in Harnad's paper is his conception of internal and external. He uses such words to refer to physical events internal or external to the brain as if the brain boundaries were some kind of relevant threshold (the magic membrane). In this regard, the debate between an internalist and an externalist ontological perspective has not yet been settled and may offers some opportunity to shed a new light on the ontological foundation of the problem considered by Harnad. Furthermore, assuming that the mind is indeed internal to anything may be a misleading conceptual move. We have no definitive evidence that the mind is inside the skull (Tonneau 2004; Rockwell 2005; Noë 2009; Manzotti forthcoming, 2011).

Once Alfred North Whitehead wrote that, "When you are criticising the philosophy of an epoch, do not chiefly direct your attention to those intellectual positions which its exponents felt it necessary explicitly to defend. There will be some fundamental assumptions which adherents of all the variant systems within the epoch unconsciously presuppose. Such assumptions appear so obvious that people do not know that they are assuming because no other way of putting things has ever occurred to them. With these assumptions a certain limited number of types of philosophic system are possible, and this group of systems constitutes the philosophy of the epoch." (Whitehead 1925, p. 48) Similarly we do not want to criticize Harnad's arguments – they appear to be sound and inescapable once a certain set of assumptions is taken for granted. Yet, this set of assumptions may be the biggest obstacle to the understanding of the relation between doing and feeling.

References

Manzotti, R., (forthcoming, 2011), "The Spread Mind. Is Consciousness Situated?" in Teorema.

Noë, A., (2009), Out of the Head. Why you are not your brain., Cambridge (Mass), MIT Press.

Rockwell, T., (2005), Neither ghost nor brain, Cambridge (Mass), MIT Press.

Strawson, G., (2011), Soul Dust. The Observer.

Tonneau, F., (2004), "Consciousness Outside the Head" in Behavior and Philosophy, 32: 97–123.

Whitehead, A. N., (1925), Science and the modern world, New York, Free Press.

Functional Explanation is Causal Explanation

(Reply to [Antonio Chella & Riccardo Manzotti](#))

Stevan Harnad

Antonio Chella & Riccardo Manzotti suggest that since we know that feeling exists, any explanation that cannot account for it is inadequate. They also suggest that there is a difference between functional explanation and causal explanation, illustrating the difference with examples from physics. Functional explanation may not explain feeling, but causal explanation may succeed, perhaps partly by scrapping the distinction between states that are internal and external to the brain:

CHELLA & MANZOTTI: *"since the fact that we feel is an empirical[ly] undeniable fact albeit from a first-person perspective, we should argue against any view that does not predict such possibility."*

Except if no causal theory can explain feeling -- in which case we are better off with one that can at least explain doing than with no explanation at all.

CHELLA & MANZOTTI: *"If feeling [does] not fit into the functional description of reality, so much the worse for functionalism."*

So much the worse for any causal explanation. The Turing Robot is "merely" indistinguishable from is in performance capacity, but the Turing biorobot also has equivalent internal processes and states, even if synthetic ones. That's still normal causal explanation, and remains so even if the biodynamics are natural rather than synthetic.

In other words, there is no wedge to be driven between "functional" explanation and "causal" explanation: All dynamical explanations of feeling are equally ineffectual, for the same reasons: There is neither any causal room for feeling, nor is there any causal need for them.

CHELLA & MANZOTTI: *"we purposefully shifted from a causal description to a functional one"*

But unfortunately it is a distinction that marks nothing substantive, and does not solve the "hard" problem of explaining how and why we feel.

CHELLA & MANZOTTI: *"the equations for gravity and electromagnetism have the same form... The two cases are functionally identical. Yet, they are different both in causal and in physical terms since the physical properties (or powers) which are responsible for the two situations are very different (on one hand, mass and gravity and, on the other hand, electric charge and electromagnetic force)"*

The equations are equivalent at one level of description, but they are not a complete description. Both mass and charge are measurable, describable, predictable physical properties -- unlike feelings, which certainly exist, but do

not otherwise enter into the causal matrix.

CHELLA & MANZOTTI: *"What is still missing is a theory outlining a conceptual and causal connection between neural activity and phenomenal experience and functionalism does not seem to possess the resources to do it."*

Nor does any other causal theory.

CHELLA & MANZOTTI: *"[In] Harnad's... conception... internal and external... refer to physical events internal or external to the brain as if the brain boundaries were some kind of relevant threshold..."*

Yes, mental states (feelings) -- for which I recommend a migraine headache as a paradigmatic example -- occur in the head, not outside it. Both doings and their functional substrate can be distributed beyond the bounds of a head, but feelings (until further notice) cannot...

For a critique of the notion of the "extended mind," see:

Dror, I. and Harnad, S. (2009) [Offloading Cognition onto Cognitive Technology](#). In Dror, I. and Harnad, S. (Eds) (2009): *Cognition Distributed: How Cognitive Technology Extends Our Minds*. Amsterdam: John Benjamins

CHELLA & MANZOTTI: *"assuming that the mind is indeed internal to anything may be a misleading"*

It is misleading to mix up "in the head" with "in the mind." But "mind" is a weasel word. To have a mind is to feel. And there is no reason to doubt that a headache cannot be wider than a head...

