

Review of: Wegner, Daniel (2003) *The Illusion of Conscious Will*. MIT Press.

Multiply reviewed in: *Behavioral and Brain Sciences* (2004) 27: 649-692

Precis: <http://www.bbsonline.org/Preprints/Wegner-05012003/Referees/Wegner.pdf>

On the Incommensurability of Feeling and Doing: The Illusion of Free Will

Stevan Harnad

Chaire de recherche du Canada
Institut des sciences cognitives
Université du Québec à Montréal
Montréal, Québec Canada H3C 3P8
<http://www.crsc.uqam.ca/>

and

Department of Electronics & Computer Science
University of Southampton
Highfield, Southampton, UK SO17 1BJ
<http://www.ecs.soton.ac.uk/~harnad/>

ABSTRACT: Wegner's book is on perceptual and motor illusions of free will and agency, but free will is much more profoundly illusory than any of the phenomena reported in his book. We never do anything because we feel like it, even it feels as if we do, because feelings cannot have any independent causal power, on pain of telekinetic dualism.

The difference between a conscious entity and a nonconscious entity is that the conscious entity feels and the nonconscious one does not. What the feeler feels is just a detail. That it feels is the critical property, and explaining how and why it feels is the so-called "hard" problem of explaining consciousness.

Among the kinds of feelings a feeler might feel are primary sensations (such as colors, shapes, sounds, smells, touches), pains, emotions (such as joy, fear), desires (such as hunger, lust), kinesthesias (involuntary movements, voluntary movements), and complex cognitive states (such as knowing that it is raining, believing that it is Tuesday, understanding that $2+2=4$, and meaning that the cat is on the mat). These are all just things a feeler might be able to feel.

There is also a world, to which those feelings may somehow be related. When a feeler feels that it is getting colder, the temperature may or may not be dropping; if it feels it has moved its limb, the limb may or may not have moved (if it has a limb at all). And if it feels it has moved its limb voluntarily, it may or may not have been voluntary: It may have moved because a physician tapped a reflex point or even just pushed the limb, and the feeler mistakenly felt it as a voluntary movement.

So there are feelers and feelings and a world in which the feelings take place, with what is felt sometimes corresponding to what is actually happening in the world, sometimes not. Let us pause to think about this "correspondence": I said the feeler could feel it is

getting colder, and the temperature (internal or external) might indeed be dropping, or it might not. When the feeler feels it's getting colder, and it really is getting colder, we are tempted to say that it has felt the drop in temperature: but a drop in temperature is, as we all know, a reduction in the average motion of the molecules in the (say) ambient air molecules. But does what "feeling colder" feels like really resemble in any way a reduction of average molecular motion? Does what red looks like resemble a property of reflected photons of a certain frequency? Does what palpating a round shape feels like resemble a property of roundness (equidistance from a center, or symmetric curvature)? Does even what it feels like for something to feel more or less intense resemble what it is for for any physical magnitude to be greater or less in any other way than that it is reliably correlated with it?

I didn't really want to embark on an extended Lockean exercise here, just to pump some intuitions about the relation between feelings and their causes and correlates. Our feelings must correlate with reality closely enough to get us by in this Darwinian world, but that is a functional criterion and not a geometric one. We know what it means for a drawing of a circle to resemble a circle, but what does it mean to say that what it feels like to look at that drawing resembles the circle (as opposed to resembling what it feels like to look at the circle)?

So in what way might our feeling of free will be an illusion? We know there are sensory illusions as well as hallucinations: I can see something when there is nothing there. I can see one line as longer when they are both the same length. In such cases we would say that the usually reliable correlation between the way things really are in the world and the way they feel has failed us. It does not, of course, mean that all perception is an illusion. Just that it can sometimes be illusory, i.e., that our feelings can sometimes mislead us.

Wegner's book reviews many examples of how our sense of agency, voluntariness and control can also fail us, creating what one might call kinesiological illusions (kinesthetic illusions being already an inhabited niche). Under some conditions we feel we are voluntarily doing something that we are not, or are not voluntarily doing something that we are. The effect sometimes even stretches from one feeler to another, with one feeler having the illusion that they are willing the movements of the other feeler, or vice versa. Here too, the natural conclusion is that there can be illusions of will and voluntary control, including all the usual perceptual possibilities: false negatives, false positives, and misses. It does not, of course, follow that free will is an illusion. Just that our feelings can sometimes mislead us.

Yet of course free will is in another sense – a sense other than the usually reliable correlation between the way things feel and the way things are – an illusion. When I make a voluntary movement such as raising my arm (and let's say this is not one of the cases when the correlation fails and creates an illusion), it feels as if I am causing my arm to raise: I've raised it because I felt like it. But could it be that my feelings are no more the cause of the motion here than they resemble the motion of molecules, when I am feeling the heat or the cold? There is something more fundamental here than just the

incommensurability between physical properties and feelings. Psychophysics at least gives us a reliable quantitative correlation, if it cannot give us a qualitative resemblance. But with action we are not talking about mere correlation or resemblance, but about causation. What could it mean for feelings to be causes?

An innocent (and probably correct) meaning would be that the same neural event that causes the feelings causes the movement. But then the real cause is that neural event, not the feeling (Libet BBS). Could feelings themselves, independently, be the causes of a movement? Only if a certain kind of dualism were true: Telekinetic (“mind over matter”) dualism. But for telekinetic dualism to be true, there would have to be reliable empirical evidence that feelings can be independent causes, rather like gravitation or electromagnetic forces. Yet centuries of efforts by paranormal researchers to demonstrate that this was the case have so far failed to generate any reliable evidence (Alcock BBS) – and the evidence would have to be dramatic indeed, for it would either be in conflict with the conservation laws of physics or it would need to add mental forces into the pantheon of physical forces (of which there are otherwise currently only four). A tall order for a soft science!

In the absence of any evidence for telekinetic dualism, we are left with the “hard” problem: If feelings can’t be causes, why are they there at all, and how are they implemented, physically? (Not "What are their neural correlates?", but "How and why does neurophysiological function -- 'functed' function -- become felt function -- unlike, say, cardiac, or pulmonary, or renal function, which we 'funct' but do not feel or feel we cause?") Wegner suggests that “constructing a virtual agent” somehow has some functional utility. If it did have some functional utility, then free will would be elevated from the correlational to the causal, and the hard problem would be solved. But is it solved?

It is easy to provide a plausible adaptive function for the capacity to detect and respond to input, and to learn to interact differentially with different kinds of input, via sensorimotor interactions of both the voluntary and involuntary kind. (The difference between pushing and being pushed.) You could even put a social world into this, with conspecifics, kin, allies, competitors, predators, and even natural language communication. A correlation between sensory activity and reality would be useful and adaptive, as would a correlation between motor activity and reality, including the all-important distinction between the causes and effects of when the organism does something voluntarily, owing to an internal command, as opposed to when it does something involuntarily, owing to some inner automatism or an external force. Correlations would not be restricted to inert ones: Correlations between outer behavior and inner states could evolve and be learned so that such organisms could communicate as well as “mind-read” – not in the true sense of telepathy, but simply through adaptive functional correlations.

There could be mistakes (“illusions”) of both the sensory and motor sort, in which the usually reliable functional correlations fail. But so far one element is missing from this rather lifelike and humanlike scenario: Feelings. How and why should these organisms be feelers? It will not do to propose, as Wegner does, that it is functionally useful to

“construct a virtual agent”: It is certainly useful to detect information, learn patterns and correlations, plan, infer, reason, compute, simulate, even simulate actions and the actions of others. But those are merely functions, and functions we can already implement with inert pieces of man-made machinery. Why and how should any of those very useful and adaptive functions – right up to language – become felt functions, rather than just function? For unless we can explain how and why those functions are felt functions, all the rest of our attempt to explain how and why we do things because we feel like it comes to nothing at all.

FOOTNOTE on "correlation": The observation that feelings "correlate quantitatively" with reality even if they do not resemble it qualitatively" is a cheat. The right construal of Wittgenstein's argument against the possibility of inventing a "private language" is that it does not even make sense to say that my sensations correlate quantitatively with reality: In fact, it only *feels (qualitatively) as if* they correlate quantitatively with reality! The objective correlation is only between reality and what I do, not what I feel. That feeling I have now, that this sound is more intense than it seems to have been a second ago, and that this correlates (objectively and quantitatively) with the increased amplitude registered by the audiometer (and is the same correlated feeling of increased intensity that I have felt and have confirmed with meter readings countless times before) is, objectively speaking, merely a correlation between the sound and what I (and my neurons) do. Whether it is also in some way "correlated" with what I feel neither I nor anyone else (nor instruments) can determine. It just feels that way. Feeling and function are not only causally dissociated, they are objectively incommensurable.

Notes.

The book is not about the illusion of will -- conscious” is redundant: “unconscious will” is self-contradictory, just as an “unfelt feeling” would be – but about illusions of will: misperceived and misattributed agency and causality

Mostly about illusions or misattributions of causality: cf hallucination. This doesn't imply that FW is an illusion any more than sensory illusions imply that sensation is illusory. It is logic (and metaphysics) that shows FW is an illusions. We don't actually do things because we feel like it; it only feels like it!

Sense of agency can be altered or duped. But that does not make it an illusion: all perception can be duped. Some times we perceive correctly, sometimes incorrectly, whether about the length of two lines or the voluntariness of a movement. But the fact that no movement is really voluntary at all is another matter.

The basic problem is the causal status of feeling (of which the feeling of agency, or voluntariness, is just one of an infinity of the same kind.

Hysteric automatism can come from hypnosis, which can suppress pain too, but is it like

neurological automatic writing? (anosognosia)

The only one who believes in free will is the telekineticist (immaterial force, mind over matter). Rest of us are just telling teleological Just-So Stories, post/ad hoc hermeneutics, not explanation

Anosognosia for the causes of our own behavior.

Libet, Alcock, BBS

P 328 P 340: errors

Social Psychology (dissonance) of rationalization and confabulation
Not about reasons why but about feeling that

Not FW vs. determinism but the causal status of feeling: feeling willing is its phenomenological core: It feels as if I am doing things because I feel like it. Not why we feel will, but why we feel at all.

Can't separate problem of consciousness and causation. Sensorimotor feelings

“action projection” is false, but mind-reading is true.

Telepathy, Telekinesis and Teleology: Causality and Consciousness: Feeling and Function: Explanation

Why one me? There are no unfelt feelings, and we can't really feel another's feelings (it just feels-as-if feeling another's feelings, i.e., it's an illusion)

Cogito: I might feel-as-if I am another, or many, or nothing, but clear-headed reflection (w/o anosognosia) shows this is not so: It is infallible that feeling is felt. And feeling of feeling is part of the package.

Robot “morality”: feelings only help because they are correlated with objective adaptive “values” – values are still underdetermined.

“People experience conscious will when they interpret their own thoughts as the cause of their action” interpret? Or feel as if they are doing something because they feel like it

not like billiard-ball to ball causation because there objective cause is inferred, whereas subjective will is felt. Causation may sometimes be inferred because we are hard-wired to infer it: “It looks (and feels) as if A causes B, but I might be wrong.”

I might also be wrong that something that feels voluntary really is voluntary, but I can't be wrong that it feels voluntary.

“The consistency principle” How can a thought resemble, or relate to, or “be consistent

with” an action: It can only feel-as-if it resembles, etc. But that’s the same as feeling as if it causes. Hence it is no explanation.

FW sense is not the same thing as an illusion of control. FW is just the illusion of having free will at all, whereas control may be an illusion in this case and not in that case. (cf “illusion of seeming”)

That new ideas do not feel willed goes against the notion that trying in advance makes the outcome feel willed. Do we will the answer to $2+2=?$ Do we will adverse thoughts? Remembering, or thinking an old thought, is every bit as sudden and spontaneous as a “creative” thought... (attributed to “unconscious mind”: does that have a FW too?)

“Construction of a ‘virtual agent’” To be conscious one needs neither to be able to do, nor to feel one is doing nor feel that one is doing voluntarily: one need only feel.

Speculating about functional (= causal) role of FW is as futile as speculating about the functional (causal) role of C (feeling). Why feel anything at all (including, in particular, feeling as if one is causing what one is doing)?