Inferences are just folk psychology

Thomas Metzinger

Philosophisches Seminar, Johannes Gutenberg-Universität Mainz, D-55099 Mainz, Germany. **metzinger@uni-mainz.de**
http://www.philosophie.uni-mainz.de/metzinger/

Abstract: To speak of "inferences," "interpretations," and so forth is just folk psychology. It creates new homunculi, and it is also implausible from a purely phenomenological perspective. Phenomenal volition must be described in the conceptual framework of an empirically plausible theory of mental representation. It is a non sequitur to conclude from dissociability that the functional properties determining phenomenal volition *never* make a causal contribution.

I have offered an alternative interpretation of some of Dan Wegner's most relevant data elsewhere (Metzinger 2003, p. 506ff), and will confine myself to three conceptual points here. Wegner's project could be further strengthened by eliminating an omnipresent version of the *mereological fallacy,* by adopting an empirically plausible *theory of mental representation,* and by avoiding certain kinds of non sequiturs.

In his laudable attempt to describe and more carefully analyze the functional architecture of phenomenal volition Wegner frequently employs personal-level concepts and predicates like "interpreting" (e.g., thoughts as causes [Wegner 2002, p. 64ff]), "inference" (e.g., of an apparent causal path, p. 68ff.), or "control" (e.g., mental control, p. 310ff). The author uses such predicates and concepts simultaneously on personal and subpersonal levels of description. At one time he speaks of the whole person as interpreting, for instance, her own thoughts as causes, and at another time of an "interpretive system" on the subpersonal level (e.g., as a course-sensing mechanism, p. 317); at one time he analyzes the person as a whole exerting mental control, at another he talks about a "controlling apparatus" (e.g., p. 312), and so forth. The subpersonal readings are all fallacious: Brains – or functional subsystems of brains – don't interpret anything, they don't make any inferences, and they don't exert control. Only whole persons can be directed at the meaning of certain sentences (or of sentences describing chains of internal events), thereby attempting to interpret them. Only whole persons could establish inferences between mentally represented propositions. And, only whole persons can be directed at the fulfillment conditions defining certain goal-states, that is, only whole persons can truly make an attempt at controlling a certain state of affairs.

The deeper problem in the background is that one needs an empirically plausible and conceptually coherent theory of mental representation to successfully describe the architecture of phenomenal volition on a subpersonal level, that is, without committing the homunculus fallacy. Daniel Wegner does not develop such a theory, but assumes that apparent mental causation results from "interpretations" and "inferences." Brains, however, are not inference machines, but associative engines (see, e.g., Clark 1989; 1993). Probably brains are even more than that, namely, complex dynamical systems exhibiting something like a "liquid" architecture. It has now become overwhelmingly plausible that such systems do not exhibit a critical property which Ramsey et al. (1991) have called "propositional modularity"; the fact that they represent propositional content in a way that makes individual units functionally discrete, semantically interpretable, and endowed with a distinct causal role. In this light the "inferences" underlying apparent mental causation are a leftover piece of folk psychology that has to be substituted by a suitable subsymbolic/dynamical story. Second, "inferences" and "interpretations" also are phenomenologically implausible, because none of us actually subjectively *experience* themselves as drawing inferences and interpreting syntactical structures before having the conscious experience of will. They are leftover pieces of folk phenomenology. As a matter of fact these two points can now be seen as a new constraint for all candidate theories of mental representation: Are they able to accommodate a fine-grained and subsymbolic analysis for the architecture of conscious volition, functionally as well as phenomenologically?

Every form of phenomenal content has at least one minimally sufficient neural correlate (Chalmers 2000). This is true of every instance of consciously experienced volition too: For every such experience there will be a minimal set of neurofunctional properties that reliably activates it and which has no proper subset that would have the same effect. Many philosophers would even argue that every single instance of phenomenal volition is token-identical to this very correlate.

Interestingly, in a given system, every single overt action has at least one such minimally sufficient neural correlate too. For every such action there will be a minimal set of neurofunctional properties that reliably brings it about, and which has no proper subset that would have the same effect. Dan Wegner has made a major contribution in showing how many situations there are in which behavior and phenomenal will can be *dissociated* in various ways, and what the parameters guiding such dissociations are. Given his data, it is a rational and plausible conclusion to assume that both kinds of sets of neurofunctional properties only loosely overlap. At times they can be instantiated in isolation. What does *not* follow is the proposition that – especially in nonpathological standard configurations – the functional properties determining phenomenal volition never make a considerable contribution to action control. This is a non sequitur.

What we have to distinguish are cases where apparent mental causation is *mere appearance,* and cases where appearance and mentally represented knowledge possibly coexist. In philosopher's jargon, we need a criterion that allows us to distinguish between those cases when conscious will is only phenomenal content, and cases where epistemic, intentional content is co-instantiated in the very same event. Let me give an example: In standard configurations the functional properties determining the fact that the experience of conscious will occurs could at the same time be a subset of exactly those functional properties that make the self-organizing dynamics of a certain, ongoing motor selection process globally available, thereby adding flexibility, context-sensitivity, integration with working and autobiographical memory, availability for attentional processing, and so forth. The "feeling" of will could then be not an illusion, but, rather, a nonconceptual form of self-knowledge – that is, the introspective knowledge that one right now is a system undergoing the internal transformation just described.

Differentiating dissociation and repression

John Morton

Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom. **j.morton@ucl.ac.uk**

Abstract: Now that consciousness is thoroughly out of the way, we can focus more precisely on the kinds of things that can happen underneath. A contrast can be made between *dissociation* and *repression.* Dissociation is where a memory record or set of autobiographical memory records cannot be retrieved; repression is where there is retrieval of a record but, because of the current task specification, the contents of the record, though entering into current processing, are not allowed into consciousness. I look at hypnotic amnesia and dissociative identity disorder in relation to this contrast.

Wegner has set up a framework within which phenomena such as post-hypnotic amnesia and *dissociative identity disorder* (DID) sit very comfortably, even though the paradoxes are heightened. Consider, if conscious will is an illusion, then acting without the experience of consciously willing one's actions can be seen as realism.

I will focus on amnesia in the contexts of hypnosis and DID. As a cognitive psychologist, I am interested in the nature of the amnesia. I regard the autobiographical memory system as separate

from a central cognitive processor and separate from the buffer store that services this processor. I also consider that the material in the buffer is not automatically accessible to consciousness. Such a position sits well with the framework that Wegner has so elegantly laid out.

There are two major ways in which a person may fail to consciously retrieve an autobiographical memory. The first kind of problem is that the memory record in question cannot be accessed. There are a number of ways in which this could happen (cf. Mortonet al. 1985), but the outcome is that the material in the record does not arrive in the buffer store. It cannot, then, in this model, influence behaviour. The second kind of problem is that the material arrives in the buffer store, will be subject to some processing, and can influence behaviour. However, some mechanism exists that prevents this material being made available to conscious processing.

How might one distinguish these two broad classes? Consider the following experiment: You perform a free association task with a participant. Then, after some manipulation or other, you perform the identical free association task, using exactly the same stimulus words in the same order. What will be the influences on the outcome the second time through? Roughly speaking, there will be two influences. The first will be the priming of the response given the first time around due to activation remaining in the perceptuo-semantic-motor system. Let me call this *implicit priming,* for short. The second will be the memory record of the first run through the task. Now, suppose that the experimental manipulation involves a hypnotic suggestion that the first task be forgotten. In principle, the forgetting might be achieved in either of the two ways outlined above. If the record of the first task is inaccessible, then its only influence would be that of the implicit priming. There should be a lot of repeat responses and they should be faster than in the first run. In fact, when the study was run (Mortonet al. 2000) hypnotised subjects took longer on their second run than on their first run (N=14; mean response time first run, 1.4 seconds; second run, 1.9 seconds). They also produced a mean of 7.9/15 different responses on the second run. These figures compare with a control group who were encouraged to give the same responses who were faster on the second run (1.18 sec vs. 1.1 sec; 1.7/15 repeat responses). The decrement in the hypnotised group could only have come about if they had retrieved the information concerning the first run and used it to monitor their responses during the second run. In other words, they had repressed the information.

Consider the same experiment run with participants diagnosed with DID. The alters of interest are those who claim, directly or indirectly, that they have no knowledge of the activities of the other alters. You run the free association experiment with one alter, switch to a second alter who claims to know nothing of what has just happened and run the experiment again. If the second alter has no access to the memory records laid down by the first alter – what I call dissociation – then (individual differences between the alters permitting) the second alter should be faster than the first. I have run this experiment on three DID patients to date. With DID1 the response times were 1.7 secs for alter 1 and 1.5 secs for alter 2 in one experiment and 1.4 to 1.2 secs in a second experiment. With DID2 the times were 3.1 to 2.7 seconds. These preliminary data indicate that these DID patients are dissociating – in contrast to exhibiting repression like the hypnotised subjects. Finally, DID3 showed an increase in response time from 1.7 to 1.9 seconds. This patient had failed to give clear evidence of dissociation in other tests, and so the association task seems to have some diagnostic utility.

Wegner (2002) supposes that DID switching is equivalent to rebooting a computer with a different operating system (p. 269). My own feeling is that DID switching is more like logging out and then relogging in under a different user name, with a denial of access to the personal files of the other users, though with the same operating system and user programs. Otherwise we would not find priming.

Finally, I have been wondering whether it is possible to ask what book Wegner actually intended to write. It could not have been more fun than this one.

Free will and the varieties of affective and conative selves

Jaak Panksepp

Department of Psychology, J. P. Scott Center for Neuroscience Mind and Behavior, Bowling Green State University, Bowling Green, OH 43403; and Department of Biomedical Engineering, Falk Center for Molecular Therapeutics, McCormick School of Engineering, Northwestern University, Evanston, IL 60208. **jpankse@bgnet.bgsu.edu**

Abstract: A causally efficacious conscious will is a small part of our everyday activities, but a part that deserves to be recognized, studied, and cherished, perhaps as a fundamental, emotion- and conation-related, right hemispheric neuronal process. Such brain functions might be less in doubt if we consider all the pieces of the larger pie, especially those where our passions and desires reside.

Wegner (2002) offers fascinating journeys through carnival aspects of the human cognitive apparatus. He gracefully coaxes us to abandon a cherished belief: that our actions revolve around willfulness within the conscious universe of the brain/mind. As long recognized, "the whole subject of *unconscious cerebration* . . . is pregnant with interest" even as some "draw what must be regarded as untenable and artificial distinctions between reality and *resemblance* in conscious and unconscious mental action. They suggest, if they do not assert, that purposive actions may possess a false appearance of ideation, a deceptive volition" (Lindsay 1879, p. 7). Wegner proceeds steadily in that direction, with modest conviction.

Should mind scientists finally agree that human thoughts cannot voluntarily intend actions? Not at all, if only a modest slice of pie is presented as the whole, especially since our left hemisphere "interpreter" is so commonly a "confabulator" (Turnbull & Solms 2004).

Many credible scenarios are left. What about the measured actions and potentially "ironic" willfulness of our self-absorbed and pessimistic right-hemispheric "realist" (Davidson & Hugspeth 1995)? What about the consciousness of those ancient emotional operating systems that generate our animalian intentions-in-action (Panksepp 2003a; 2003b)? Although the extroverted left hemisphere enjoys a good story and pontificates obsessively to grease the social wheels, might other brain areas be more adept at provoking self-consciously motivated actions?

Free will may be more critically linked to imagery-attuned functions of the right hemisphere, in close touch with periconscious subcortical emotional functions. Wegner does peer behind William James's "flimsiest of screens" as he exhibits menageries of peculiar mental proclivities. However, he avoids our deeper animalian nature, wherein persistent desires and willfulness are not just social constructions, but animalian attributes of our dopamine fired *seeking* urges (Panksepp 1998a). The feeling of *conation,* resurrected briefly as a willful "cognitive emotion," is a promising candidate from that periconscious realm. Such intention-generating processes are not deeply unconscious, although they often fade under glaring Hollywood-like screens of perceptual and linguistic consciousness.

Here is my recent encounter with the pure feeling of free will: During surgery under spinal blockade, I could no longer feel myself voluntarily wiggle my feet, but move they did, predictably, verifiable by looking. That spooky feeling of efference, without somatosensory/proprioceptive feedback, was part of my volitional apparatus. Might the periconscious *conative* borderland between our animalistic "intentions in action" (affective consciousness) and our human "intentions to act" (cognitive consciousness) be where