Philosophie de l'esprit

Is there metaphysical free will?

Probably not. There is a strong argument that we are token-identical with something physical and it seems that the "space" in our physical world does not suit sufficiently our intuitions about free will since we face the following dilemma: in a deterministic world, we are not the ultimate origin of our acts and therefore hardly free; this is the consequence argument against compatibilist positions. On the other hand, it seems problematic to base our free will on indeterministic processes since it then could not be distinguished from chance; this is the matter of chance argument.

Still, this paper spells out a possible solution to the matter of chance argument: 1) I discuss the notion of autonomy of special science's properties *within* a conservative reductionist framework and thereby some kind of independency and robustness of psychological properties. Against this background, 2), it is possible to focus in more detail on the possible adaptive functions of noise / indeterministic brain processes if combined with particular constraints and/or feedback mechanisms in the brain. The upshot of this consideration is a possible distinction between free will and chance *within* indeterminism *but without* contradicting neither ontological nor epistemological reductionism.

It is common ground that biological property tokens are identical with something physical (cf. Kim 2005). However, biological property *types* can be multiply realized by different configurations of physical properties. This is to say that property tokens of one biological type (B) can come under different physical types (P_1 and P_2). Hence, there are biological similarities (brought out by B) among property tokens in the world despite of their physical differences. However, as it has been shown recently (cf. Esfeld & Sachse 2011, ch. 5), this asymmetry does not imply an anti-reductionist position since the physical differences between tokens coming under P_1 and P_2 may have an impact on the biological level: by changing the environmental conditions, functional differences between the tokens coming under P_1 and P_2 can be observed from a biological point of view as well. Therefore it is possible to construct in biological terms functionally defined sub-types B_1 and B_2 of B in such a way that these sub-types are nomologically co-extenstional with P_1 and P_2. We can then reduce *conservatively* B by means of its sub-types to physics, which, as a *theoretical* possibility seems to be necessary in order to avoid conflicts with both ontological reductionism and the completeness of physics.

The upshot of this strategy is that it opens a new perspective on biological similarities, in particular the sensitivity and robustness of biological processes to physical variation (that may be due to indeterministic processes, low number effect, thermal agitation, etc.). As mentioned before, the conservative reductionism framework is chosen here to avoid possible conflicts with ontological reductionism and the completeness of physics for our following consideration of "noise" (as a possible placeholder for indeterministic processes in the brain) and neurobiological constraints and feedback mechanisms. In order to establish the link to the free will debate, let me start with an analogy.

Biological evolution is in the last resort based on gene mutations that may lead to different phenotypic effects that are either selectively neutral of positively/negatively selected. The latter case is commonly known "differential reproduction" by natural selection. This selection process is the fundamental mechanism of evolution and the cited conservative reduction model is applied here for comparing neutralism and selectionism (cf. Nei 2005):

under normal standard environmental conditions, there are physical differences between genes possible that do not lead to differential reproduction (multiple realization; physical differences that are selectively neutral). However, since the environmental conditions continuously change, it is just a question of time and specific environmental conditions when the physical differences in question have an impact on differential reproduction and thus lead to biological evolution (selection of physical differences that were selectively neutral beforehand; cf. Rosenberg 1994, p. 32). Hence, depending on the given environmental conditions, biological properties are more or less dependent on its physical composition. In other words, it is not the case that physical differences are *always* and *in the same way* causally relevant in the context of natural selection. This is the independency of biology from physics without supporting an anti-reductionist position.

The question now is whether there exists a similar independency at the psychological level that may serve us for being free but does not rely on some kind of libertarianist position (that conflicts with both ontological reductionism and the completeness claim of physics)? In what follows, we thus take for granted that our mental property tokens are token-identical with something neurobiological or, in the last resort, something physical. However, mental property types can be multiply realized in the brain. This is to say that property tokens of one mental type (M) can come under different neurobiological types (N_1 and N_2). Therefore, one can say that there are mental similarities (brought out by M) among mental property tokens in the world despite of neurobiological differences: the neurobiological differences between tokens of N_1 and N_2 may have an impact on the mental level: by changing psychological testing conditions, the neurobiological differences between tokens coming under N_1 and N_2 can be observed from a psychological point as well. Therefore it is possible to construct in psychological terms functionally defined sub-types M_1 and M_2 of M in such a way that these sub-types are nomologically co-extensional with N_1 and N_2. It thus seems that we can reduce *conservatively* M by means of its sub-types to neurobiology (this reflection is based on the cited reduction model applied to the philosophy of mind; see Soom & Sachse & Esfeld 2010).

Against the background of this general framework, the same perspective as for biology opens up on psychological similarities, and thus psychological kinds, laws and explanations. It is thus not the case that neurobiological differences are *always* and *in the same way* relevant for psychological properties. However, neurobiological differences may play a similar crucial role for our free will as physical differences of genes do for the fitness of the organism in question and thus evolution. The uncontroversial reflection for what follows is that, by positive and negative feedback mechanism, our brain is constantly changing, selecting, and thus adapting/learning. Oversimplified, mental reflections and followed executed actions are subject to negative and positive feedback mechanisms as can be shown very easily in conditioning experiences. To give you now a preliminary idea of the more central idea of this paper, let us hypothetically imagine that, e.g. in our brain, there sometimes occur indeterministic biological processes. Indeterminism is here understood as objective ontological indeterminism according to certain interpretations of quantum physics that affect under certain conditions the neurobiological processes. Since there is no empirical question on *whether* they affect neurobiological processes but only *when* and *how often* does this happen in a human brain, we can continue our discussion of how the brain may "treat" changes that are possibly due to indeterministic processes.

For instance, let us schematically consider a neuronal signal transmission between two neurons, say *a* and *b* that has some positive (at least non-negative) effect for the organism. Take the case that someone tells you to do X (and this input activates neuron *a* in your brain), and then you do X (since the neuron *a* activates your neuron *b* that leads to do action X). Thus, the transmission from *a* to *b* is not blocked and usually takes place whenever *a* is activated (someone tells you to do X). Let us now imagine that it happens again that someone

tells to do X and thus your neuron *a* is activated – but the signal transmission to *b* does not take place because of some indeterministic process (e.g. noise interference or something else). This means somehow that, by chance, you are not doing X. I do not want to claim, as others may do, that this per se makes you a free person, since up to this step, there is no real difference to some chancy event. More important, the mentionned non-signal-transmission may be by itself (due to its effects/non-effects) subject to positive and negative feedback mechanisms. Simplified, it is possible that the indeterministically occurring interruption of a signal transmission has a higher positive effect for you (than the effect of the signal transmission and always doing X when some tells you to do). Due to neural plasticity, one may show, the *non*-signal-transmission from *a* to *b* may thus become at least more probable as it was before. You may then more often than before react to input as "please do X" by not doing X. Even though indeterminism may play a crucial role in the *beginning* of our story, it is by far not the end. Keeping this general consideration in mind, I now discuss several crucial notions in more detail.

Taking for granted that there are in fact indeterministic processes in our world including the brain. As mentioned, it is then just a question of time and particular conditions that indeterministically occurring changes in the brain lead to salient neurobiological and thus psychological effects that are therefore subject to feedback mechanisms in the brain (as oversimplified presented before). It is thus, first of all, not the case that everything we do and think is determined by the past and the laws of nature. Therefore, the standard objection against compatibilist positions, the so-called *consequence argument*, does not apply at this point (cf. Kane 2002, especially part III). Second, it is possible to distinguish free will from chance even though indeterministic processes may play a crucial role to call a will free. My proposal claims only that indeterministic processes may sometimes affect brain changes that are by its effects subject to feedback mechanisms. In biology, the occurrence of a genetic mutation that leads to a phenotypic difference is in the long run a necessary premise for evolution by natural selection. However, the explanatory power of evolutionary biology is provided by the mechanism of natural selection and *not* by the occurrence of mutations per se. The same applies to the indeterministic brain processes and feedback mechanisms. Indeterministic processes are necessary in order to avoid the consequence argument. However, whether or not indeterministically occurring changes in brain lead in the long run to different reasoning, behaviour, etc. depend on neurobiological constraints (e.g. sensibility to minor physical changes) and feedback mechanisms. This is where the explanatory power comes from. This is why there is a difference between chance and our reasoning and behaviour.

Let me sum up by putting this idea in some other, still abstract and theoretical, words. Imagine that we think since years about how to develop a stronger notion of free will than it is outlined in these nice compatibilist positions without taking some libertarian move that is just hardly coherent with the common ground in the philosophy of science. We further detect salient similarities between debates proper to the philosophy of science and the debate on free will. And then, by some indeterministic process in the brain, we once think differently about free will and this new thinking suits us, get a positive feedback since it is a *metaphysical* free will, our moral reflections can thus be based on it, etc. Isn't this the beginning of some nice starting point? This is in fact where we can get the "space" for having a free will within a physical world out of conservative reductionism. As I said, there is no biological evolution just because of chance but because of natural selection of occurring gene differences. Following Gerald Edelman on Darwin, "variation in a population is not just noise but the substrate for selection" (Edelman, 2006, p. 25). In analogy, we are not free by chance per se but by some sort of selection due to feedback mechanisms of, among others, *indeterministically* occurring changes. This approach based in the first place on metaphysical

and theoretical considerations and possibilities becomes more compelling when combined with empirical work on stochastic fluctuations and its constraints and regulation on the genetic, developmental and neurobiological level (e.g. Gerhart & Kirschner 2007). It thus seems empirically plausible to link the capacity of living systems to produce (facilitated phenotypic) variation that could be at the origin of innovations and adaptations with the contemporary debate on free, thereby avoiding metaphysically problematic positions (like emergentism and libertarianism).

Edelman, Gerald (2006): *Second nature. Brain science and human knowledge*, New Haven: Yael University Press.

Esfeld, Michael & Sachse, Christian (2007): *Conservative reductionism*, New York / London: Routledge.

Gerhart, John and Kirschner, Marc (2007): "The theory of facilitated variation", *PNAS* 104, pp. 8582-8589.

Kane, Robert (2002): *The Oxford Handbook of Free Will*, Oxford: Oxford University Press.

Kim, Jaegwon (2005): *Physicalism, or something near enough*, Princeton: Princeton University Press.

Nei, Masatoshi (2005): "Selectionism and neutralism in molecular evolution", *Molecular Biology and Evolution* 22, pp. 2318-2342.

Rosenberg, Alexander (1994): *Instrumental biology or the disunity of science*, Chicago: University of Chicago Press.

Soom, Patrice, Sachse, Christian and Esfeld, Michael (2010): "Psycho-neural reduction through functional sub-types", *Journal of Consciousness Studies* 17, pp. 7-26.