Commentary

Neurophilosophy of Free Will

Henrik Walter

MIT Press, Cambridge (MA), 2001

Lorenzo Del Savio *

lorenzo.delsavio@ifom-ieo-campus.it

Neurophilosophy of Free Will frames the analytic debate about free will within current neurophysiological theories. The introductory chapter overviews several decades of discussion by listing three intuitions that should be accounted for by any eligible theory of free choice: freedom (ability to do otherwise), intelligibility (acting for reasons) and agency (being the source of our own choices). A moderate neurophilosophical *manifesto* is then outlined in the second chapter: knowledge of the brain should inform philosophy of mind. Chapter three eventually tries to meet these analytical and methodological *desiderata*: freedom, intelligibility and agency are extensively (though tentatively) naturalized by means of neuroscientific insights. Walter draws the conclusion that libertarianism should be rejected and free will explained by *natural autonomy*, a concept that should save *phenomena* and intuitions alike.

The relevance of Henrik Walter's book goes well beyond the issue that it explicitly addresses. His naturalization effort covers a wide range of traditional topics, from intentional content to the concept of a person. Many scientific theories that had been picked up were admittedly fairly hypothetical (p. 259) and so they still are. Thus the fate of Walter's specific proposals is open to scientific scrutiny. His main methodological point is nonetheless irreversible. Nobody would deny that philosophizing should be *conscious* of scientific developments. Walter claims it should be also *involved* in empirical inquiries: he urges for a *«bridge discipline between subjective experience, philosophical theorizing and empirical research»* (p. 125).

This review will focus on Walter's way of fulfilling his naturalistic program and, therefore, the first two chapters are let aside. It is nonetheless worth to

* University of Milan; IEO European Institute of Oncology

remember that, when the book was published, neither an extensive overview of the free will debate (chapter one) nor a plea for non-reductive physicalism (minimal neurophilosophy – chapter two) was yet commonsense. After more than a decade, Walter's three core proposals of naturalization instead deserve our attention. Part (1) presents Walter's ideas about chaos theory and free choices, part (2) deals with his naturalistic conception of brain content and part (3) outlines the link envisaged by Walter between the concept of a person and some neuroscientific insights about emotions.

(1) Walter's thoughts about freedom are organized in two sections: a *pars destruens* in which he argues that quantum physics is not relevant for the free will debate and a *pars construens* that borrows from chaos theory in order to dissolve the puzzle of freedom. The latter goes as follows: free choices require that, at some instant, more than one future is possible. Choices are bifurcating paths. Now, either the world is deterministic or it is not. If the former is the case, then there are no alternative paths by definition (van Inwagen, 1986) and it only seems there are. On the other hand, if indeterminism is true, then there are alternative paths at some point, but the choice is indeterminate. Therefore, which path is taken does not depend on anybody's decision (agency intuition) nor can it be explained by reasons (intelligibility intuition). The three core intuitions cannot hold simultaneously.

Penrose famously proposed to link agency and indeterministic phenomena of quantum mechanics to discard the second horn of the dilemma. According to Walter, linking agency and quantum phenomena has an obvious hurdle: agency is *prima facie* an organism level phenomenon, thus macroscopic, and macro-systems are practically deterministic. In fact, mainstream neurophysiologists take atomic and subatomic processes for granted and typically work on macromolecules (actually, dynamics of a huge number of them). To bridge the gap with the atomic level we would need an *amplifier* theory like Penrose's. A common criticism to this proposal points out that it rests on promissory notes about future physical theories. Walter's objection is rather that it is not even compatible with what we know about the brain: he gives compelling reasons to the effect that the brain is quite unaffected by atomic phenomena (p. 161).[1] Instead, we should focus on brain level phenomena rather than cell-level interactions.

[1] Those reasons are also independently interesting for who is concerned with inter-level reduction and levels of mechanistic explanations (see Darden 2008).

Compatibilist strategies often rest on a shift of meaning. Possibility of doing otherwise in the same circumstances is weakened and becomes possibility of doing otherwise *if one had wanted to*. Same circumstances are not identical, they are fairly similar – a move that resembles Lewis' thesis on identity across possible worlds. Would these counterpart-circumstances rescue freedom? Walter claims they would and try to explain why these circumstances are neurophilosophically relevant. In a nutshell, he suggests that brain network dynamics is likely to be chaotic and hence extremely sensitive to small fluctuations of parameters. «Using chaotic behavior, a cognitive system retains the option of reacting quickly, flexibly, and sensitively to relevant stimuli, changes in the environment, or ideas» (p. 182). These outcomes cannot be predicted despite their being wholly deterministic: epistemic indeterminacy suffices to account for the intuition that we could have done otherwise.

Yet an objection easily comes up: dependence on chaotic outcomes would end up in auto-epistemic indetermination. We would be astonished by our own decisions all the time. Not so, according to Walter: he takes a revisionist stance on decisions to blur the objection. «Decisions are not processes that occur at a point in time, they are events extended through time» (p. 183), indeed we would speak of mere reflexes – not choices – beneath of a certain time threshold (p. 184). Throughout the decision process, our cognitive system follows unpredictable trajectories, eventually resting down to a stable state: this is the decision. Nonetheless a major trouble remains: who controls the values that determine the trajectories? Walter admits to be again in a thicket, but a very different one indeed: the problem has shifted from availability of alternatives to agency. Walter's solution falls or stands with his naturalization of agency, a topic that is tackled towards the end of the book.

(2) Although the success of Walter's conception of free will depends mainly on later paragraphs, his treatment of intelligibility has several far-reaching philosophical consequences. The issue is spelled out in terms of acting for reasons and the latter is linked with the debate about intentionality. Millikan's ideas are then borrowed in order to explain intentionality of mental states in a physical world (neurosemantics). Intentionality is naturalized by adaptation and adaptations are explained by natural selection. Walter's step forward deals with the last concept: the scope of Millikan's teleosemantics gets wider to become *neurosemantics*. The explanatory power of selection is

stretched beyond the usual evolutionary time spans: selective-like processes in ontogenetic or even instantaneous times produce meaning *in the brain*.

Walter proposes that a physical state comes to be about a chunk of reality having a relational proper function. In Millikan's words, it is an intentional state that has been selected for conveying certain contents. Walter's specific contribution takes natural selection as a rather abstract schema and suggests that it applies to evolutionary times, ontogenetic times and even ultra-fast instants *phenomena*. The adaptive immune system is often the main example of a selective-like mechanism that produces specificity (antigens *recognition* – an intentional metaphor indeed) in ontogenetic times. The same token could well be true for the brain:

> among the constraints that support stability [of a brain structure] are not only complementary effects within the brain, but also interaction with the external world. [...] A temporarily stable state can be interpreted semantically because the stabilizing process is an adaptation. (p. 228)

An early proponent of this theory was not by chance Edelman (i.e., Edelman 1992), a Nobel-awarded immunologist.

A further dimension taken into account is subsequently the inter-subjective language, by mean of which content plays a physical causal role in the world. Causal networks including contents are *bona fide* physical interactions, nonetheless they might be *paraphrased* (p. 240) by reason talk in virtue of the selective history of their components. These intentional states, in the wording of Walter, supervene on physical structures and environmental surroundings.

Something of our intuitions about intelligibility has faded away (p. 243): an intentional state does not have causal power *as* intentional state but only as physical state, nonetheless it can be given an intentional content because of its history. It is noteworthy that the puzzle of free will vanishes even when one accepts this conclusion: here, reasons determine course of action only in a loose sense and the underlying physical process might well be indeterministic.

(3) Free choices belong to the physical (chaotic) causal network (freedom naturalized) and can be interpreted as reasons in virtue of the proper function of some physical state of the brain (intelligibility naturalized). Yet only a small subset of these reasons are recognized by a person as her own. «A compatibilist theory of agency must postulate that the determinants converging in a person are action of that person. In other words, it must be a theory about what makes

an executing instance a "self" or "person"» (p. 263). How can we naturalistically make sense of attributions of reasons to persons?

Frankfurt (1971) argued that identification with second order volitions is crucial: a person's will is free only if he is free to have the will he wants. Mention of freedom in the *definiens* would end up in a regress, but wholehearted identification with a volition guarantees that a second order will is authentically expression of a person. Yet we are not given further clarifications: Walter suggests neuroscience can provide some fruitful insights: indeed he claims that emotions play a pivotal role in this identification.

According to Damasio's work, while pondering, we simulate a counterfactual scenario by means of an imagined outcome of a choice and a correspondent body state representation. Crucially, body state representations get stabilized throughout the life of an individual, thus implicitly containing the past history of a person (p. 284). Frankfurt's regress of always higher-level volitions is stopped by emotional identification with a self-representation, Walter's naturalistic rephrasing of wholeheartedness. Only those volitions that are embedded in this emotional way are authentic. «Self-determined behavior is not a result of rational considerations, instead we learn to make clever and socially responsible decisions with the aid of our emotions» (p. 290).

Walter concludes summarizing his theory of natural autonomy:

> under very similar circumstances we could also do other than we actually do (because of the chaotic nature of our brain). This choice is understandable (intelligible – it is determined by past events, by immediate adaptation processes in the brain, and partially by our linguistically formed environment), and it is authentic (when through reflections loops with emotional adjustment we can identify with that action). (p. 299)

Whether this natural autonomy is compatibilist or hard-determinist won't concern us here: it all depends on our attitudes toward libertarianism (Kane 2001).

Still Walter's conclusions are open to conceptual as well as empirical challenges. Conceptually, the natural philosophy side of the debate loses its strength in Walter's treatment because there is no longer the issue determinism *versus* indeterminism in the foreground: natural autonomy is compatible with both metaphysics. This deflationary result instead brings authorship and the notion of agency at the core of the philosophical concern about freedom. Two main topics are therefore worth pointing out: consciousness and moral responsibility.

Although it is mentioned throughout the book, there is no extensive treatment of consciousness. Later works on free will turned to the topic claiming that consciousness lies at the very center of our conception of agency (and *hence* responsibility), namely Wegner's idea of consciousness as *emotion of authorship* (2002). Despite its frankly hard-deterministic framework, Wegner's proposal begins where natural autonomy ends. It shows how emotions, self and consciousness are entrenched.

A second topic is moral responsibility. Walter declares not to deal with moral theorizing because he wants to single out the pure metaphysical nucleus of the debate. I have argued that Walter's results turned out to be deflationary exactly as far as natural philosophy is concerned. Yet it is arguably not possible to clarify the concept of an agent without any link to responsibility, if not to explain why we have such a concept in the first place.

Aristotle's sea battle argument rested on logical worries. Medieval work on free will was carried out against a theological background. In modern times, the debate has shifted into a mechanistic framework and, perhaps more surprisingly, it has resurfaced even when statistical social laws have been discovered (Hacking 1990). Cognitive sciences have also been used as the last scenarios of this ancient battle (e.g., Libet 1985). Walter conceives the role of neurosciences more broadly. They do not simply challenge the pre-theoretical concept of free will. Rather, neurosciences might cast light on the notion of a person who is the author of her own decisions.

REFERENCES

Churchland, P. (2007). Neurophilosophy: The early years and new directions. *Functional Neurology, 22*(4), 185-195.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958-969.

Edelman, G. (1992). *Bright Air, Brilliant Fire.* New York: Basic Books.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*, 11-25.

Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press.

Kane, R. (2001). *The Oxford Handbook of Free Will*. New York: Oxford University Press.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 529-566.

Van Inwagen, P. (1986). *An Essay on Free Will*. Oxford: Clarendon Press.

Walter, H. (2001). *Neurophilosophy of Free Will: From Libertarian Illusion to a Concept of Natural Autonomy*. Cambridge, MA: MIT Press.

Wegner, D (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.