



ELSEVIER

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com



Biologie et devenir technologique de l'homme/Biology and the technological future of man

The metaphysical lessons of synthetic biology and neuroscience

*Les leçons métaphysiques de la biologie de synthèse et des neurosciences*

Bernard Baertschi

IEH2 (Institut Éthique Histoire humanités), université de Genève, 1, rue Michel-Servet, 1211 Genève 4, Switzerland

ARTICLE INFO

Available online 4 July 2015

Keywords:

Nature of life
Mind-body problem
Personhood
Moral status
Free will
Determinism

Mots clés :

Nature du vivant
Problème corps-esprit
Personnalité
Statut moral
Volonté libre
Déterminisme

ABSTRACT

In this paper, I examine some important metaphysical lessons that are often presented as derived from two new scientific disciplines: synthetic biology and neuroscience. I analyse four of them: the nature of life, the existence of a soul (the mind-body problem), personhood, and free will. Many caveats are in order, and each 'advance' or each case should be assessed for itself. I conclude that a main lesson can nevertheless be learned: in conjunction with modern science, neuroscience and synthetic biology allow us to enrich old metaphysical debates, to deepen and even renew them. In particular, it becomes less and less plausible to consider life, mind, person, and agency as non-natural or non-physical entities.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Dans ce texte, j'examine quelques leçons métaphysiques importantes, qui sont souvent présentées comme des résultats de deux nouvelles disciplines scientifiques: la biologie de synthèse et les neurosciences. J'en analyse quatre: la nature de la vie, l'existence de l'âme (le problème de l'âme et du corps), la notion de personne et la question du libre arbitre. Il est nécessaire de procéder avec précaution, et chaque « avancée » ou chaque cas doit être évalué pour lui-même. Je conclus en affirmant qu'une leçon commune peut néanmoins être tirée : en conjonction avec ce que nous apprend la science moderne, les neurosciences et la biologie de synthèse nous permettent d'enrichir ces anciens débats métaphysiques, de les approfondir, et même de les renouveler. En particulier, il devient de moins en moins plausible de considérer la vie, l'esprit, la personne et son activité morale comme des entités non naturelles ou non physiques.

© 2015 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Science and philosophy are related in a complex manner. From its beginning, in the 17th century, modern science has been seen as an alternative to classical

philosophy, that is, to medieval Aristotelianism. In his famous book, *The Origin of Forms and Qualities*, Robert Boyle argues thoroughly in order to show that hylomorphism is false and should be replaced by a conception where physical bodies are a bundle of moving particles instead of a compound of matter and form [1]. The controversy was raging and, in the end, modern science won. But philosophy was not dead, because the scientific victory

Email address: bernard.baertschi@unige.ch.<http://dx.doi.org/10.1016/j.crvi.2015.06.002>

1631-0691/© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

concerned mainly natural philosophy: epistemology, ethics, and metaphysics survived and have even flourished since then.

The rivalry between science and philosophy has not ended, but is only sporadic. Usually, it takes the form of a proposal to replace philosophy or some part of it by a scientific discipline. Think of the heroic stance of Auguste Comte in the 19th century, hoping that humanity would soon reach a scientific era, after having passed through a religious and a metaphysical one [2], and of the modest conception of scientific progress suggested by Karl Popper: 'To obtain a picture or model of this quasi-inductive evolution of science, the various ideas and hypotheses might be visualized as particles suspended in a fluid. Testable science is the precipitation of these particles at the bottom of the vessel: they settle down in layers [...]. As the result of this process ideas previously floating in higher metaphysical regions may sometimes be reached by the growth of science, and thus make contact with it, and settle' [3, p. 277]. With time, science replaces philosophy, but we have no proof that it will discard all philosophy, not even that it will be always possible.

In my opinion, Popper's suggestion is historically correct and a fruitful programme to pursue. However, it is not without pitfalls. The main one is that we are generally a bit hasty and jump to conclusions that are not warranted by the current state of knowledge. The risk is particularly salient with new sciences, because philosophically-minded scientists and scientifically minded philosophers are often enthusiastic and draw conclusions that go far ahead of their premises.

In this paper, I will examine some important metaphysical lessons that are often presented as derived from two new scientific disciplines: synthetic biology and neuroscience. As we will see, many caveats are in order, and each 'advance' or each case should be assessed for itself. I will analyse four of them: the nature of life, the existence of a soul (the mind-body problem), personhood, and free will. The first comes under the jurisdiction of synthetic biology, whereas the others are linked with neuroscience.

2. The nature of life

After the Craig Venter Institute successfully transplanted the *Mycoplasma micoides* genome in *Mycoplama capricolum* in March 2010, Roberta Kwok said that 'the team has fielded criticism for calling the resulting cell "synthetic" when the genome was essentially a replica of a natural genome and required an existing recipient cell. Hutchison [a member of Venter's team] argues that "synthetic" simply means "chemically synthesized", not newly designed' [4, p. 25]. Words matter and using expressions like 'synthetic' instead of 'replicated' is not without symbolic and philosophical impact. Another expression used, 'newly designed', also has connotations of its own. To design is not exactly the same as to create, but it is not very far away, and creation refers to a godlike action. Playing God or adopting a demiurgic stance are objections often voiced against biotechnologies by certain opponents. Even people not committed to a particular

worldview are using expressions with a theological tone in relation to synthetic biology. For instance, contrasting two ways of producing artificial living beings, one consisting of modifying the genome of existing beings and another consisting of building them from inanimate molecules, Joachim Boldt and Oliver Müller speak of 'creation *ex existendo*' and 'creation *ex nihilo*' [5, p. 388].

Theological expressions and hints abounded at the time of the Craig Venter Institute's success, and the Vatican was not the last to deny that the transplantation was a genuine creation. In the *Wall Street Journal* we find an echo of the debate, in a paper written by James DeGiulio [6], where we read quotes from *L'Osservatore Romano*, the official newspaper of the Vatican, saying that the transplantation was not life's creation, because DNA is only an engine in the service of life and not life itself. Therefore, through the transplantation, Venter's team has merely 'replaced one [of the] motors' of life. Craig Venter was nevertheless of the same mind and in another paper published in the same newspaper, co-authored with Daniel Gibson [7], he wrote: 'Kornberg did not create life in a test tube, nor did we create life from scratch. We transformed existing life into new life. We also did not design and build a new chromosome from nothing. Rather, using only digitized information, we synthesized a modified version of the naturally occurring *Mycoplasma mycoides* genome. The result is not an "artificial" life form.' (Arthur Kornberg was the first scientist who duplicated the DNA of a virus, in 1967) The words used could be misleading: strictly speaking, no human being is able and will ever be able to build a chromosome *from nothing* or to create life *from scratch* – only a God could do it. But maybe in the future a human being will be able to build a chromosome *from inorganic molecules* or to create life *from non-living elements*?

What does it mean for an organism to live? And what is required from scientists in order to create a living organism? For centuries, life has been tied with some special principle, material or not. For animism, a being is living if it possesses a non-material soul: a dualistic approach; for vitalism, it must have inside itself a principle of life, reducible to matter or not [8, (pp. 12–14)]. In an age of Darwinism, no scientist still accepts such views; they are non-scientific and if science has a metaphysical impact, they constitute metaphysical mistakes. For instance, 'No non-physical substance or force is distinctive of all instances of life', says Mark Bedau [9, p. 334]. Life is now considered as emerging or supervening from inanimate matter, through a long process of changes resulting in the advent of new properties, that exactly constitutes life. What is on the list of these properties is still in debate, but the ones most often mentioned are auto-organization, autonomy, capacity to adapt, reproduction, growth, evolution, and metabolism. The debate extends to the question whether these properties are each necessary or not, and if some of them are sufficient or not. Here, I have no need to acquire a firm opinion on this debate. It suffices for me that all the people involved in it accept the same basic assumption: life is an emergent phenomenon and is characterised by a set of properties; therefore, a living organism is a being that possesses some definite properties.

According to such a view, what does it mean to create an organism? What is required from scientists to do it? The answer is very simple: to create an organism is to produce an entity that possesses the properties characteristic of life. This requirement is nevertheless wanting, because it makes life's creation very easy for everyone, in procreation for example. Adding that the living being must be created through biotechnological means will not be enough, because cloning and medically assisted procreation would, in consequence, be methods to create life. What we ought to add is the requirement that a living being is created only if the elements from which it comes are not living. However, this is still not adequate, because no part of a living being is itself living, *in the sense* that it does not possess the properties of the whole organism. An organ, a cell, or a stretch of DNA, are not living in this sense. Therefore, we could say that Venter's team has created life, even if they deny it and if we agree with it. Consequently, we must add this qualification: the non-living elements should not come from a living organism. It must not be something like a functional biologic part. This said, it follows that if a biologist succeeds in designing a being that possesses these properties from inanimate or chemical materials, he will have created a living organism, that is, more abstractly and grandiloquently, he will have created life. Is it possible to do it?

Yes, because the modern conception of the nature of life puts no insuperable obstacle on the path to it, since life is a bundle of emergent properties, as said before. But it was not the case with scholastics [10, chap. 5]. René Descartes and Robert Boyle were eager to stress the novelty of the modern conception at the level of physics and they formulated two important theses in this aim:

- there exists an ontological identity between what is natural and what is artificial;
- there exists only one kind of change in nature.

Boyle is particularly clear on these points. In favour of the first thesis, he mentions the discovery that glass produced in a volcano's chimney and in a glassmaker's furnace is exactly the same glass, and he comments: 'I know not why all the productions of the fire made by chymists should be looked upon as not natural, but artificial bodies; since the fire, which is the grand agent in these changes, doth not, by being employed by the chymist, cease to be and to work as a natural agent: and since nature herself doth, by the help of the fire, sometimes afford us the like productions that the alchymist's art presents us' [1, p. 51]. The two types of glass are the same glass, because their stuff and the fire that produces them are the same. The fact that nature (volcanoes) or human beings are at the origin of the process is physically irrelevant in itself; it has no impact on the intrinsic properties of the product.

The second thesis is also defended by Boyle. For him and physicists contemporaneous with him, physical reality consists exclusively in moving corpuscles. Nature and human beings have consequently *only one way* at their disposal to change reality: to act on the local motion of

corpuscles. Local motion is, accordingly, the only kind of change in nature, the other kinds (generation, growth, and qualitative change) being reducible to it. Following this conception, transmutation from lead to gold becomes possible and is hoped for by Boyle: 'I could not see any impossibility in the nature of the thing, that one kind of metal should be transmuted into another (that being in effect no more than that one parcel of the universal matter, wherein all bodies agree, may have a texture produced in it like the texture of some other parcel of the matter common to them both)' [1, p. 94].

The conception of the material world that is beneath these theses is still with us today, but it was new then. In particular, for Aristotelian science, there was an ontological difference between natural and artificial beings, because only natural beings possess a nature, that is, an internal principle of change [11, p. 155]. Conversely, artificial bodies can only be acted on from outside; they have no 'spontaneity'. For Boyle, this Aristotelian thesis is profoundly mistaken: the properties of a being are independent of the way this being is created or produced, by human beings or by nature. In the case of glass – Boyle's example – its properties are the same, be it natural or artificial, and since human beings have the capacity to create artificial beings identical with natural ones, the human creation of artificial life is in principle possible.

Synthetic biology's programme could be successful given our metaphysical conception of the nature of life. However, this new discipline will not change our conception of it, since it has been in charge from the 17th century, but is and, hopefully, will continue to be a *confirmation* of it – it is its metaphysical lesson.

3. The existence of a soul (the mind-body problem)

Paul Wolpe observes: 'The claim that neuroscience itself can tell us something about the existence or non-existence of the soul is precisely the kind of claim that makes religious thinkers nervous' [12, p. 292]. Could neuroscience decide the mind-body problem or, at least, bring into the debate decisive facts or arguments? If religious thinkers become nervous, it is of course because they fear that the facts and arguments would be in favour of a materialist position. Is their fear justified? In order to be able to address seriously and precisely the question, I will, as in the case of the nature of life, make some historical comments and, as before, go back to the rise of modern science, because here too, the question was newly framed in this period [13].

As is well-known, René Descartes offers a dualist conception of the human being, following his criticism of Scholastics. The body and the mind – or the soul – are two substances and, as such, are distinct and separable, whereas in the Aristotelian tradition, they were two principles (form and matter), even if the Christian theologians added that the soul is also a kind of substance. For Descartes, the distinction of the mind and the body is a direct consequence of the *cogito*. As he says in the *Principles of philosophy*, the *cogito* is 'the best way to discover what sort of thing the mind is, and how it differs from the body [*l'âme est une substance entièrement distincte du corps*]' [14,

p. 95]. Then, he tries to explain their union, which is an empirical fact for him, but still needs to have its mechanism understood. 'Nature likewise teaches me by these sensations of pain, hunger, thirst, etc., that I am not only lodged in my body as a pilot in a vessel, but that I am besides so intimately conjoined, and as it were intermixed with it, that my mind and body compose a certain unity' [15, p. 492], he says, insisting that the soul is situated in the pineal gland, where it is able to act on the brain and to receive impressions from it, through the animal spirits that circulate in the nerves.

If it remains difficult for Descartes to explain in detail the mechanism through which the mind-body union is realised, its possibility has been quickly doubted. Thomas Hobbes was the first to criticise it, but Princess Elisabeth formulated the most powerful objection: in her opinion, postulating the existence of relations between the mind and the body is problematic, because it would violate one of the fundamental principles of modern physics: the principle of the conservation of the quantity of movement ($m \cdot v$). When a body is moved, the movement it acquires comes from the body that is moving it, and this last body loses exactly the same amount of movement that the first gains. Therefore, if we make abstraction of any friction, the total amount of movement will remain constant. However, the soul is necessarily motionless, since it is not a body; consequently, it cannot pass on a movement to the body. The only possibility then is that the soul creates some movement, but this movement would come *ex nihilo* and it will increase the total amount of movement present in the world, violating the fundamental principle that Descartes himself has established.

Descartes will be unable to reply correctly to this challenge and, in order to do so, his followers will abandon either the interaction, or the hope of finding a solution. John Locke will adopt this latter stance in being agnostic on the mind-body problem: 'It [is], in respect of our notions, not much more remote from our comprehension to conceive that God can, if he pleases, superadd to matter a faculty of thinking, than that he should superadd to it another substance with a faculty of thinking [16, p. 908]. Materialist monism and dualism are sent back to back, since we are deprived of the necessary cognitive capacities to decide between them. Contrariwise, Leibniz and Malebranche still defend a dualist position, but without any interaction between soul and body. Instead, they claim that the two substances live in complete parallelism: what is happening in the body corresponds to what is happening in the soul, but without any causal link. Leibniz illustrates the three dualist positions (interactionism, occasionalism, and pre-established harmony) in the following manner: 'Imagine two clocks or watches which are in perfect agreement. Now this can happen in *three ways*. The *first* is that of a natural influence [...]. The second way of making two clocks, even poor ones, agree is always to assign a skilled craftsman to them who adjusts them and constantly sets them in agreement. The *third* way is to construct these two timepieces at the beginning with such skill and accuracy that one can be assured of their subsequent agreement' [17, p. 500]. The craftsman or the constructor is, of course, God.

In the 18th century, many authors follow one of these positions, but a new manner to solve the conundrum comes on stage: rejection of dualism in favour of materialist monism. As we saw, Locke had contemplated this possibility; Spinoza had defended a variant of it, but a specifically modern version was only proposed with La Mettrie and Diderot. They claim that thought, a property of the mind, is of the same nature as sensibility, a property of every living being explaining the movements an animal makes and the perception it has of its environment. But an animal has no immaterial – and immortal – soul. Before Lamarck, materialists hesitate if sensibility, the proto-psychic property of living beings, resides in all (living) matter or supervenes on it, along with the growing complexity characteristic of some organisms (i.e. is an emerging property), but the principles of this doctrine are already clearly formulated. Diderot asks: 'What is *sensibility*? Is it a general property of matter? Is it a property resulting from organic structure?' [18, p. 109]

Pierre-Jean-Georges Cabanis, one of the great physicians during the French Revolution and an outstanding member of the *Idéologie*, summarises the materialist solution with the help of a famous metaphor: the brain is an organ specially designed to produce thought, as stomach and bowel produce digestion, liver bile and salivary glands saliva. Impressions activate the brain as food activates the stomach. So we can say with confidence that the brain, so to speak, digests the impressions, thought being the organic secretion of the brain [*le cerveau digère en quelque sorte les impressions, il fait organiquement la sécrétion de la pensée*] [19, p. 138]. In October 1838, Charles Darwin writes in his *Notebook*: 'Love of the deity, effect of *organisation*, oh you materialist! ... Why is a thought being a *secretion* of brain, more wonderful than gravity a property of matter? It is our arrogance, our admiration of ourselves' [20, pp. 340–341]. Thought is a product of organisation, i.e. of the special union of physical particles that constitute the living organism of each human being, and especially of one part of him: the brain.

This conception will become the dominant one in the scientific community. It is not surprising, since it is in agreement with the modern conception of the world and scientists are no longer attracted by very complex and counterintuitive metaphysical doctrines, except perhaps some physicists. With one famous exception: John Eccles, a dualist neuroscientist who was looking for a kind of action of the mind on the brain devoid of any energy transfer [21, chap. E7].

Moreover, materialism is supported by several conceptual and empirical arguments.

Concerning the conceptual arguments, one of them is, in my opinion, decisive. It is the *causal exclusion argument* put forward by Jaegwon Kim. Take two brain events (B and B^*) and two correlative mental events (M and M^*), related by causal relations, figured by full and dotted arrows (Fig. 1).

Brain events cause other brain events (full horizontal arrow) and their corresponding mental events (full vertical arrows), but these mental events are also causally active on other mental events and on some brain events (dotted arrows). For example, the perception of a danger (M),

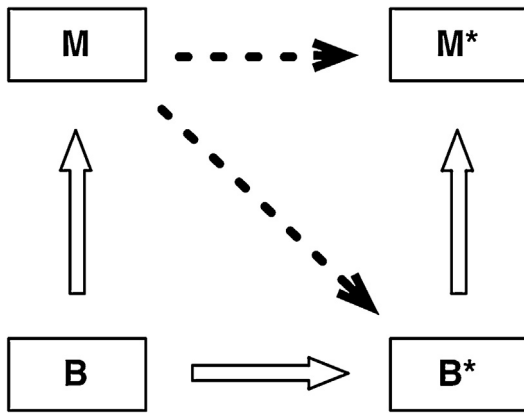


Fig. 1. Causal relations between brain and mental events.

caused by an activity in the visual cortex (B), causes some amygdala activation (B^*), and then a feeling of fear (M^*). However, it is easy to show that in such a situation only some causal relations are real (the full arrows), the others being only spurious ones (the dotted arrows), because they are redundant: the full arrows are sufficient to explain all that needs to be. It follows that mental events have no independent causal efficacy: ‘The M -to- M^* and M -to- B^* causal relations are only apparent, arising out of a genuine causal process from B to B^* ’ [22, p. 45].

Indeed, if B , correlated with M , causes itself B^* , correlated with M^* , M does not need to cause anything, and then does not cause anything. Consequently, M has no proper causal power and is causally reducible to B . Mental causality, as independent from brain causality, is only seeming causality: we have the impression that an idea as such causes another one, that an emotion as such causes another one, but it is a mistake: there is no horizontal causality between mental events, no descending causality from mental events to cerebral events exists, ontologically speaking. Consequently, the world is causally closed and no non-material entity can act in our physical world. This constitutes the world’s causal closure thesis, a conceptual counterpart of what Descartes had glimpsed on the physical level.

Until now, neuroscience has been completely absent, except as an historical curiosity with Cabanis’ sentence and as an illustration of the causal closure thesis. Materialism was established independently and long before our contemporary knowledge of the brain. Does it mean that neuroscience has no metaphysical lesson to offer here, and that the fear of religious thinkers is inappropriate? Not quite. First, religious thinkers should be afraid, but because of modern science itself: the dualism has been refuted for almost five centuries, if we require, as we should in my opinion and in Popper’s, that our metaphysics be compatible with the best scientific data we have. Second, because contemporary neuroscience gives many examples of the non-independence of mental events with regard to brain ones. Think of Phineas Gage’s story told by Antonio Damasio: brain damage has completely changed the worker’s personality, unfortunately for the worst [23].

Think also of Capgras condition, where a patient no longer recognises the people he lives with and believes that they have been replaced by doppelgangers. William Hirstein notes: ‘In the case of both Capgras syndrome and asomatognosia, something very close and very familiar, a loved one or a body part, is claimed to be unfamiliar. Just as some patients with asomatognosia will claim that the arm they see is someone else’s arm, the Capgras patient will claim that people who should be very familiar to him have actually been replaced by someone else’ [24, p. 482]. For Vilayanur Ramachandran, the illusion is explained by an abnormal perception of familiar persons due to a brain problem; it is abnormal, because it is not accompanied by the usual emotions we feel for such people. Consequently, the patient builds a rationalisation in considering the person he is with as an impostor [25, p. 514]. Here, the ‘soul’ is completely misled by the body, without any possibility to abstract from it and to see the truth. Moreover, it concerns the aspect and face of persons, a very important basis for morality, as we will see in the next section.

Adrian Raine gives another illuminating example. In the *Iowa gambling task*, imagined by Damasio’s colleague Antoine Bechara, a subject is placed in front of four piles of cards. He has to pick cards successively from the piles, as he wants. Piles A and B give large rewards in money, but sometimes larger losses. Piles C and D give lesser ones. After some time, it becomes evident that it is better to pick cards on piles C and D . Normal subjects model their behaviour on this discovery, whereas patients with ventral prefrontal damage [like Gage] do not, and exhaust their money. Raine comments that when subjects become aware of which piles are bad, we observe ‘a skin-conductance response (a somatic marker), a bodily alarm bell warning them that they were about to embark on a risky move. Subconsciously, their body knows that bad news is just around the corner, and that they would hold back on their response – but consciously their brain does not’ [26, p. 142]. Raine is not always scrupulous in his choice of words: he speaks of the brain instead of the mind. Maybe it is innocuous here, maybe not; however, what interests me is the cognitive role of the body and of the brain/mind. Here too, both are intermingled.

The soul is deeply rooted in the body, and the evidence is so overwhelming that the fact that we observe only correlations and not causality directly do not prevent us from thinking that what is happening in the body and in the brain is the cause of what we mentally experience. The mind has no independence. Experience and metaphysics converge on this outcome. Perception, emotion, and judgment are mental events inseparable from their brain correlates and completely sensitive to any modification in the latter: to change the latter is to change the former. In the 18th century, similar observations had already been made and interpreted in a materialist way. La Mettrie, for example, said: ‘What was needed to change the bravery of Caius Julius, Seneca, or Petronius into cowardice or faintheartedness? Merely an obstruction in the spleen, in the liver, an impediment in the portal vein. Why? Because the imagination is obstructed along with the viscera, and this gives rise to all the singular phenomena of hysteria

and hypochondria' [27, p. 10]. Consequently, this judgement La Mettrie passed on human beings remains inspiring: 'These proud and vain beings, more distinguished by their pride than by the name of men however much they may wish to exalt themselves, are at bottom only animals and machines which, though upright, go on all fours' [27, p. 110].

4. Personhood

The concept of 'personhood' or of 'what to be a person consists of' has been a central tenet in moral philosophy or ethics for a long time. A person, stated Boethius in the 5th century, is an individual endowed with reason (or rationality). Aquinas, Locke, and Kant have adopted this definition, and underlined that we have the strongest moral duties towards such entities [28, chap. 4]. However, in a paper entitled 'Personhood and Neuroscience: Naturalizing or Nihilating?' published in 2007, Martha Farah and Andrea Heberlein have argued that, from a neuroscientific point of view, this concept should be dispensed with. They think that its neural basis consists of an automatic response from an innate brain module or network, which is dedicated to face recognition and therefore is alien to morality (and often even to reality – they speak of an 'illusion'). For these authors, we would be well advised to turn to a conception of moral duties grounded in interests, a conception akin to utilitarianism [29]. Here, the metaphysical lesson of neuroscience extends to morality: we have resorted to a conception of morally important beings grounded on their rationality (another name for the soul?), but we were mistaken.

What is, more precisely, Farah and Heberlein's argument?

Etymologically, 'person' comes from the Latin 'persona' and from the Greek 'prosopon'. In origin, these words refer to a theatrical mask, and before a face ('what is before the eyes of somebody'). It is in this sense that neuroscientists use it when they speak of 'prosopagnosia', for a condition where a patient no longer recognises the faces of the people around him, a condition caused by a lesion of their infero-temporal cortex. Prosopagnosia is a very disturbing condition, psychologically and socially, because face recognition is central for our relations with the people we live with. For Farah and Heberlein, the psychological phenomenon of face recognition is linked with the ethics of personhood, that is, with the crucial role the moral status of person plays in most of our ethical approaches. An ethical view I will name 'personism', following Jean-Yves Goffi [30].

For Farah and Heberlein, if to be a person is so important for us, it is because face recognition is innate and very soon comes into play: 'Evidence for the innateness of the person – non-person distinction comes from the behaviour of new-born infants. Johnson et al. showed that new-borns tested within 30 minutes of birth show a greater tendency to track moving face-like patterns with their eyes than other patterns of comparable complexity or symmetry. This finding implies that, prior to virtually any opportunity to learn, the human brain is equipped with a general representation of the appearance

of the human face' [29, p. 43]. To buttress their claim, they still mention the case of a boy 'who sustained visual cortical damage, including damage to the fusiform face area, in his first day of postnatal life. Despite his relatively preserved ability to recognize non-face objects, he never acquired the ability to recognize faces.'

At first sight, it seems that these observations do not prove anything: the boy does not recognise faces, but he knows that he is before persons. However, Farah and Heberlein think that cases like this allow us to conclude that it is on the basis of our capacity to recognise faces that we have built our view of moral status. Briefly said, we are so wired that we separate spontaneously and naturally persons and non-persons, and give this distinction an ethical impact.

This is a piece of psychology. But what should we think of this natural tendency ethically speaking? Is its moral impact justified? A personist will not be very happy with that: every human being possesses a face, but every human being is not a person, because every human being is not endowed with rationality – think of some PVS patients or of anencephalic babies. Moreover, every genetically human being does not possess a face – think of embryos – and there exist non-human beings who are persons, already said Boethius, for example, angels and God. Engelhardt mentioned ET in the same context [31, p. 107]. Farah and Heberlein also deem that this natural tendency can lead us astray, when they state: 'The human face is a powerful trigger cue that activates the whole person network, and this may be what makes it hard for many of us to dismiss the personhood of a vegetative patient or a foetus' [29, p. 45]. Our innate capacity to recognise faces is obviously a source of confusion, because it compels us to grant the moral status of a person to everyone possessing a face, and this criterion is not more appropriate than, for instance, species belonging ('speciesism' is a charge often voiced against such criteria).

Farah and Heberlein extend the critique against personism *as such*, and they give three reasons for that: the concept of a person rests on an illusion, it is arbitrary, and it is a categorical concept, ill-suited to the gradual character of our psychological and moral life.

4.1. The charge of illusion

'The first relevant feature of the person network in the brain is its separateness from the systems representing other things. We suggest that this feature is responsible for the illusion that persons and non-persons are fundamentally different kinds of things in the world, despite our inability to draw a principled line between them. This illusion may come from the operation of two separate and incommensurate systems of representation in the brain for persons and for things in general, in contrast to a common distributed representation' [29, p. 45].

Our brain divides the world in a manner that is not truly divided: the brain does not carve out the world at its joints. Consequently, concerning personhood, 'like visual illusions, it is the result of brain mechanisms that represent the world nonveridically under certain circumstances.' Of course, there exist clear cases of persons – adults, for

example – but others are not clear, and their number is growing with the progress of medicine (embryos, foetuses, comatose persons, psychopaths, ...); therefore, to believe that the world is tightly divided between persons and non-persons is a mistake, of the nature of an illusion, because it is a mistake that is not responsive to arguments or reform.

Is the charge convincing? We could already doubt that the illusory character of the distinction could be an argument against personism, in the sense that the illusion is not that persons exist, but that the divide is sharp. But that is only one aspect of the illusion that will be discussed later (it is the third charge); another is that we extend personhood to beings that are not able (or not able anymore) to have interests, like some comatose individuals, granting them a moral status they should not have. However, even if personhood is anchored in a brain module for face recognition, this link is only contingent and has for long been severed by moral philosophers adopting personism.

In order to understand the gist of this reply, we have to examine how the attribution of moral status works. When we ask someone: ‘Why do you grant moral status to X’, he usually points to some property of X. For example: ‘Human beings have moral status because they are rational beings’, or ‘Animals are morally considerable because they are sentient (i.e. they can feel pleasure and suffer)’. ‘To be rational’ or ‘to be sentient’ are properties conferring moral status, and if they confer it, it is because they are valuable. Following G. E. Moore, we will consequently say that a being’s value is the value he possesses by virtue of his valuable properties [32, p. 33; 33 (pp. 9–10)].

Consequently, we *value* persons because they possess properties bearing *values* (rationality), not for the fact they have a face. It is possible that the psychological origin of this process resides in our sensibility to faces, but this fact has no essential relation with our moral judgment, it is only contingent, and often misleading, because the class of individuals with a human face and the class of individuals endowed with reason are not the same, even though they overlap.

4.2. *The charge of arbitrariness*

Conceptually, personhood could be the marker of a moral status, as the possession of interests could be (because having interests is also a bearer of values). But is it an adequate one? Farah and Heberlein do not think it is, because it is arbitrary.

This charge is nevertheless a little ambiguous, and covers two different objections. The first is that the property of rationality is a gradual one, and that we are unable to say precisely where the threshold is. The second is that rationality covers a large number of different properties and that it is arbitrary to pick one instead of another. The authors illustrate both objections in discussing Fletcher’s view: ‘Joseph Fletcher proposes 15 criteria for personhood. He begins with intelligence, and makes an admirably straightforward effort to specify the dividing line between persons and non-persons by referring to intelligence quotient (IQ) scores: “Below IQ 40 individuals might not be persons; below IQ 20 they are definitely not

persons.” The problem with this criterion is that, while it is explicit and precise, it is also arbitrary. His other 14 “marks of personhood” include traits and capacities similar to the ones already mentioned as well as a few additions and elaborations. They are: self-awareness, self-control, sense of time, sense of futurity, sense of the past, capacity to relate to others, concern for others, communication with other persons, control of existence, curiosity, change and changeability, balance of rationality and feeling, idiosyncrasy, and neocortical function’ [29, p. 38]. The second charge (arbitrariness of the property) will be examined here, and then the first one (arbitrariness of threshold).

Should we possess conscience, consciousness, critical interests, capacity for language, for abstract thinking, ability to choose on the basis of reason, or to have a life plan in order to be a person? All these properties, and those mentioned by Fletcher, can be put under the umbrella of reason or rationality, but it seems possible to have one without the others. Traditionally, reason has been put forward in contradistinction to sentience, rationality being the mark of human beings and sentience the mark of animals. But nowadays we cannot be satisfied anymore by such a conception, since some animals possess some rational capacities listed above and some human beings do not.

This difficulty is often raised against personism. Nevertheless, it is not fatal at all: for ethics we do not need a definition of what a person is, that is a necessary and sufficient condition of personhood; we do not even need a necessary condition, but only a sufficient one, that is a criterion. And we have many, including the items on the list of mental properties put forward by Fletcher–Roskies speaks of a ‘cluster-concept’ [34, p. 56]. Practically, this means that we will require a being to possess at least one of the rational properties on the list in order to be granted the status of a person.

4.3. *The charge of graduality*

A person must possess at least one of the rational properties on the list; but these properties are gradual. Therefore what amount of it? Where is the threshold? It seems impossible to set it non-arbitrarily, even with the help of neuroscience: ‘Relevant clinical observations and neural network modelling indicate that the change in psychological capabilities would be gradual and would in general lack the kinds of qualitative transition points that could be used as non-arbitrary places to draw a line between persons and non-persons’ [29, p. 40]. Since this failure extends to all psychological properties, *including interests*, a more modest project is finally proposed: ‘The real contribution of neuroscience to understanding personhood may be in revealing not what persons are, but rather why we have the intuition that there are persons’ [29, p. 40]. And we know the answer: face recognition.

I conclude that since the moral status of human beings necessarily leans on a human property, even a psychological one, neuroscience can help us to tell what are the possible candidates and teach us why we tend to pick one rather than another; but it remains and will remain unable to tell what is the normatively adequate one.

5. Free will

In a well-known penal case in the US, a man was found guilty of child molestation, because he had sexually-forbidden sex behaviour with his daughter-in-law. Some time later, he was admitted to hospital for violent headaches and a brain tumour was discovered. After its removal, this man's behaviour returned to normal for a long time, but he reoffended and it was observed that his tumour had relapsed [26 (pp. 303–305); 35]. If this case became famous, it is because of the close connection observed between the successive states of his brain and his behaviour – a new evidence of the embodiment of our soul. However, courts of law are regularly confronted with defendants whose brain is somehow not normally functioning. In these cases, courts frequently ask neurologists and psychiatrists to determine if the defendants are responsible for their deeds or not. If they are not, or not completely, the penal verdict will be mitigated or even changed to a medical one ('for reason of insanity', as it is said).

In most jurisdictions, in order to be considered as responsible, an accused should be in normal cognitive and volitive conditions: he must be able to understand that his behaviour was bad or unlawful, and not have been coerced to act, by an external threat or an internal compulsion. It is the same with normal persons: they are responsible for their acts (*E*) if they are not seriously cognitively and volitionally impaired, that is, if they are free. In short: *A* is morally or legally responsible for *E* if *A* has intentionally caused *E* and *A* could have acted otherwise or not have acted at all. However, here the difficulties begin, because 'A could have acted otherwise' can receive two meanings, a strong and a weak one.

The weak one is the moral or legal one: *A* could have acted otherwise if he was cognitively and volitionally normal when he acted. If it is not the case, for example if he had a brain lesion like Gage, he could not and, consequently, he is not responsible. Determining responsibility and its level is 'business as usual' for lawyers and forensic psychiatry.

The strong meaning is not linked with our social practices, but with our conception of human beings as parts of the natural order of the world. In this sense, in order to be able to have acted otherwise, we must have *free will*.

What does free will consist of? Imagine you have bought a bottle of red wine, because your spouse said you will have meat for dinner. You could have bought white wine if you had chosen to, but since you prefer red wine with meat, you have chosen not to. You are completely responsible for your choice, in the weak sense. Imagine now that you return at the time just before you decided to buy red wine. Could you then have bought white wine even if all the events that have happened until your decision are completely identical with the events that have prompted you to buy red wine? If you think you could, then you claim that you have free will, that is, a power to choose independently of what happens in the world. If not, if you do not have free will, you cannot act differently and consequently are not free and responsible in the strong

sense, even if you are cognitively and volitionally normal. This is a genuine metaphysical question, contrary to freedom and responsibility in the weak sense, which can be named a psychological and ethical question.

Do we have free will? There have existed for a long time serious doubts that we have it. In the Ancient World, the belief in fate undermined the belief in free will. Diogenes Laertius reports that Zeno, the Stoic philosopher, had condemned a servant to be whipped, because he had stolen an object. As he objected that it was his destiny to steal, Zeno added: and to be whipped! [36, p. 805] In the medieval period, divine predetermination replaced fate, but the challenge for free will was the same. Following Augustine, Jean Calvin claimed that our posthumous fate is set before our birth, because God knows from immemorial time who will be saved and who will go to hell [37]. This difficulty was never overcome, but in Modern Times, predetermination was replaced by physical determinism, an upshot of modern science once more. The route from one question to the other is strikingly exposed in Leibniz's *Essai de théodicée*, and its canonical formulation is in Laplace: 'We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed [...] would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes' [38, p. 4].

Determinism is the upshot of the scientific conception of the world; therefore free will ought to be given up if metaphysics should be in agreement with science. But this renunciation has its price, because free will is linked with moral and legal responsibility. The argument goes this way:

- 1) if an agent *A* is morally responsible for an act *E*, then *A* could have refrained from doing *E*;
- 2) if *E* hangs uniquely on the previous state of the world *W*, then *E* is inescapable;
- 3) if *E* is inescapable, then *A* could not have eschewed doing *E*;
- 4) therefore *A* is responsible for *E* only if *E* does not uniquely hang on *W*;
- 5) *E* does not uniquely hang on *W* only if it additionally depends on a causal power independent of *W*;
- 6) free will is such a causal power;
- 7) therefore *A* is morally responsible for *E* only if *A* possesses free will;
- 8) free will is incompatible with the truth of physical determinism;
- 9) therefore *A* does not possess free will, and consequently *A* could not have eschewed doing *E*;
- 10) conclusion: *A* is not morally responsible for *E* (and is never responsible for any decision or act of its own).

For many authors, this conclusion is not acceptable: we cannot give up responsibility, because it is basic to all our practical institutions (notably morality and law). How then to get out of this dilemma (responsibility vs determinism)?

When two propositions p and q are contradictory, it is logically possible to reject p or to reject q . So, we can logically reject determinism or free will [39]. Those who accept the conclusion (i.e. propositions 1 to 10) reject free will; they are called *hard determinists*, a conception defended by Diderot and the *Idéologues*. Hard determinists usually think that responsibility is an illusion and that we ought to change our moral and legal practices accordingly. *Libertarians* keep free will and reject determinism: they deny propositions 9 and 10 and, consequently, must weaken the causal closure of the world, usually by appealing to quantum mechanics (Eccles, in [21]) [40], a move we have already encountered.

But hard determinism and libertarianism are not the only options. When two propositions p and q are contradictory, it is still possible to modify p or q , in order for them to become compatible. *Compatibilists* adopt this strategy, denying proposition 8. Conceding the truth of determinism, they modify the meaning of free will accordingly. In their conception, free will is no longer the power to decide and choose without being impeded *tout court*, but without being hindered by certain causes rather than others. Coercion is such a cause, so a coerced act is not free and implies no responsibility; but the previous state of the world W is not such a cause and usual acts are free, implying responsibility [41 (pp. 201–237)].

Here too, the causal closure of the world is crucial, and it is a thesis that has been established long before the growth of neuroscience. Does it mean that here we are in the same position as with the mind-body problem: neuroscience illustrates the dependence of our mind on our brain, i.e. the materialist and determinist conception of the world linked to modern science, but nothing more? This confirmation stance is already no small progress, but here, discussions have been wider, especially on two points: social consequences and voluntary movement (Libet's experiments).

Already in the 17th century, Jean Barbeyrac noted in a footnote, in his French translation of Pufendorf, that if ever we could know that all human actions were necessitated, that is devoid of free will, we should not disseminate this, lest human beings stop acting and become completely lazy [42, p. 134, n. 1]. More recently, a study has shown that when people are exposed to statements denying free will, they cheat more easily: 'Subjects who read a series of sentences, such as "Science has demonstrated that free will is an illusion," are [...] more likely to cheat and steal' [43, p. 18]. Metaphysical beliefs could have a practical impact!

More metaphysically interesting are Libet's studies. Around 1970, he conducted a series of experiments on voluntary movement and obtained puzzling results. The subjects were asked to press a button and to report the exact moment when they decided to do it. With the help of electroencephalography, Libet found that the brain begins the bodily movement some time before the subject's decision (usually about 300 ms before). Daniel Dennet comments: 'When you think you're deciding, you're actually just passively watching a sort of delayed internal videotape (the ominous 300-millisecond delay) of the real deciding that happened unconsciously in your brain quite a while before "it occurred to you" to flick'

[44, p. 229]. Some authors immediately thought that Libet had demonstrated the illusion of free will: we believe that we decide, but the brain has already begun the movement, therefore this organ is the actual decision-maker, but as it is only a physical organ, free will is an illusion. However, this interpretation has been hotly contested, by Libet first, and by many other authors since then. Even Dennet does not think that Libet's experiments are an argument in favour of his deterministic position, since he says: 'What Libet discovered was not that consciousness lags ominously behind unconscious decision, but that conscious decision-making takes time' [44, p. 238].

More recently and in the same spirit – deciding takes time – Itzhac Fried and colleagues have observed that every decision entails at least two stages: an accumulation of information, and a mechanism to stop this accumulation in order to take the decision: 'The neuronal process suggests a mechanism whereby the feeling of will arises once integration of firing of recruited medial frontal neurons crosses a threshold [...]. These findings lend support to the view that the experience of will emerges as the culmination of premotor activity (probably in combination with networks in parietal cortex) starting several hundreds of ms before awareness' [45, p. 557].

The debate goes on, but manifestly, we cannot expect that any experiment conducted according to Libet's scheme will be able to decide between libertarianism and determinism. The progress in understanding voluntary movement will nevertheless allow us to better understand human actions and agency, two very important topics for responsibility in the weak sense.

However, here too, the illustrations given by neuroscience coupled with the metaphysical lessons of modern science do not leave us in serious perplexity or doubt. No non-material entity resides in our body – materialism is true – and as long as science explains the behaviour of macro objects like our body in causal terms, determinism will be true. What we observe in our brains – neurodeterminism – is only an aspect of the universal determinism: it is physical determinism in play. As I said in a paper written with Alex Mauron: 'Neurodeterminism [...] appears to be a reformulation in modern scientific terms of the traditional doctrine of determinism *tout court* [i.e. physical determinism] when it is applied to human beings. If every event that has an influence on our behaviour can only have an effect through some brain events, if all brain events are determined by causal chains of events, and if every mental state is a brain state, then neurodeterminism is determinism *tout court* applied to human actions. Accordingly, for the first time in history, we are able to observe determinism *tout court* at work in our mental life' [46 (pp. 158–159)].

What remains to be decided, and it will not be with the aid of neuroscience, is whether the final truth resides in hard determinism or in compatibilism, also named soft determinism.

6. Conclusion

After this long journey, we could be a little disappointed: the metaphysical lessons of neuroscience and of

synthetic biology are slim. However, to feel such a negative emotion would not be appropriate. Remember Popper: science progresses *slowly* and its encroachments on metaphysics are only *partial*. But they are real: in conjunction with modern science, neuroscience and synthetic biology allow us to enrich the metaphysical debates and to renew them. It becomes less and less plausible to consider life, mind, person, and agency as non-natural entities. Corroboration is an endless process and since we are never sheltered from a falsification or a refutation, the adventure continues to be exciting, but with modern science, we are on steady metaphysical ground.

Disclosure of interest

The author declares that he has no conflict of interest concerning this article.

References

- [1] R. Boyle, *The origin of form and qualities, The works of the honourable Robert Boyle, Vol. 3*, London, 1666.
- [2] A. Comte, *Cours de philosophie positive*, Hermann, Paris, 1998.
- [3] K.R. Popper, *The logic of scientific discovery*, Routledge, London, 1992.
- [4] R. Kwock, DNA's master craftsmen, *Nature* 468 (2010) 22–25.
- [5] J. Boldt, O. Müller, Newtons of the leaves of grass, *Nat. Biotechnol.* 26 (2008) 387–389.
- [6] J. DeGiulio, Venter denies synthetic cell discovery is artificial life, *Vatican agrees*, *Wall Street J.* (2010) www.patentdocs.org-venter-denies-synthetic-cell-discovery-is-artificial-life-vatican-agrees.html.
- [7] C. Venter, D. Gibson, How we created the first synthetic cell, *Wall Street J.* (2010) www.wsj.com-SB10001424052748704026204575266460432676840.html.
- [8] Swiss federal ethics committee on non-human biotechnology synthetic biology – ethical considerations Bern, Switzerland, 2010 www.ekah.admin.ch-e-Synthetische_Bio_Broschuere.pdf.
- [9] M. Bedau, The nature of life, in: Boden (Ed.), *The philosophy of artificial life*, Oxford University Press, Oxford, 1996, pp. 332–357.
- [10] B. Baertschi, *La vie artificielle*, CENH, Bern, Switzerland, 2009.
- [11] É. Gilson, *Études sur le rôle de la pensée médiévale dans la formation du système cartésien*, Vrin, Paris, 1984.
- [12] P. Wolpe, Religious responses to neuroscientific questions, in: J. Illes (Ed.), *Neuroethics*, Oxford University Press, Oxford, UK, 2006, pp. 289–296.
- [13] B. Baertschi, *Les rapports de l'âme et du corps. Descartes, Diderot et Maine de Biran*, Vrin, Paris, 1992.
- [14] R. Descartes, *Les principes de la philosophie*, in: F. Alquié (Ed.), *Œuvres philosophiques*, Vol. III, Garnier, Paris, 1973.
- [15] R. Descartes, *Méditations métaphysiques*, in: F. Alquié (Ed.), *Œuvres philosophiques*, Vol. II, Garnier, Paris, 1973.
- [16] J. Locke, *An essay concerning human understanding*, London, 1690.
- [17] G.W. Leibniz, *Éclaircissements du système nouveau de l'âme et du corps*, in: C.I. Gerhardt (Ed.), *Die philosophischen Schriften von G.W. Leibniz*, Vol. IV, Reprint Olms, 1978 [transl. D. Scott].
- [18] D. Diderot, *Commentaire*, in: F. Hemsterhuis, G. May (Eds.), *Lettre sur l'homme et ses rapports*, New Haven, Yale University Press, 1964.
- [19] P.J.G. Cabanis, *Rapports du physique et du moral de l'homme*, Slatkine Reprints, Genève, 1980.
- [20] D. Weiss, *Toward the vindication of Friedrich Engels*, in: R.S. Cohen, M.W. Wartofsky (Eds.), *Methodology, metaphysics and the history of science*, Reidel, Dordrecht, The Netherlands, 1984, pp. 331–358.
- [21] K.R. Popper, J.C. Eccles, *The self and its brain*, Springer, Berlin, 1997.
- [22] J. Kim, *Mind in a physical world*, MIT Press, London, 1998.
- [23] A. Damasio, *Descartes' error: emotion, reason and the human brain*, Putnam's Sons, New York, 1994.
- [24] W. Hirstein, *Confabulations about people and their limbs present or absent*, in: J. Bickle (Ed.), *The Oxford handbook of philosophy and neuroscience*, OUP, Oxford, UK, 2009, pp. 473–512.
- [25] J. Mundale, S. Gallagher, *Delusional experience*, in: J. Bickle (Ed.), *The Oxford handbook of philosophy and neuroscience*, OUP, Oxford, UK, 2009, pp. 513–521.
- [26] A. Raine, *The anatomy of violence*, Penguin Books, London, 2013.
- [27] J.O. La Mettrie, *L'homme-machine, Œuvres philosophiques*, Vol. I, Fayard, Paris, 1987.
- [28] B. Baertschi, *Enquête philosophique sur la dignité*, Labor et Fides, Genève, 2005.
- [29] M. Farah, A. Heberlein, *Personhood and neuroscience: naturalizing or nihilating?* *Am. J. Bioeth.* 7 (2007) 37–48.
- [30] J.-Y. Goffi, *Personne, personnalisme, personnisme*, in: R. Mache (Ed.), *La personne dans les sociétés techniques*, L'Harmattan, Paris, 2007, pp. 13–41.
- [31] H.T. Engelhardt, *The foundations of bioethics*, Oxford University Press, Oxford, UK, 1986.
- [32] W. Rabinowicz, T. Rønnow-Rasmussen, *A distinction in value: intrinsic and for its own sake*, *Proc. Aristotelian Soc.* 100 (2000) 33–51.
- [33] B. Baertschi, *The moral status of artificial life*, *Environ. Values* 21 (2012) 5–18.
- [34] A. Roskies, *The illusion of personhood*, *Am. J. Bioeth.* 7 (2007) 55–57.
- [35] J. Burns, R. Swerdlow, *Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign*, *Arch. Neurol.* 60 (2003) 437–440.
- [36] D. Laërce, *Vies et doctrines des philosophes illustres*, Pochothèque, Paris, 1999.
- [37] J. Calvin, *Institution de la religion chrétienne*, Vol. III, Les Belles Lettres, Paris, 1939.
- [38] P.S. Laplace, *Essai philosophique sur les probabilités*, Bachelier, Paris, 1840.
- [39] R. Kane (Ed.), *The Oxford handbook of free will*, Oxford University Press, Oxford, UK, 2002.
- [40] J.R. Searle, *Freedom and neurobiology*, University Press of California, Columbia, 2008.
- [41] P. Churchland, *Brain-wise*, MIT Press, Cambridge, MA, 2002.
- [42] S. Pufendorf, *Les devoirs de l'homme et du citoyen*, Vol. I, Presses Universitaires de Caen, Caen, France, 2002.
- [43] E. Nahmias, *Why "willusionism" leads to "bad results": comments on Baumeister, Crescioni, and Alquist*, *Neuroethics* 4 (2011) 17–24.
- [44] D. Dennett, *Freedom evolves*, Allen Lane, London, 2003.
- [45] I. Fried, R. Mukamel, G. Kreiman, *Internally generated preactivation of single neurons in human medial frontal cortex predicts volition*, *Neuron* 69 (2011) 548–562.
- [46] B. Baertschi, A. Mauron, *Genetic determinism, neuronal determinism and determinism tout court*, in: J. Illes, B.J. Sahakian (Eds.), *The Oxford handbook of neuroethics*, Oxford University Press, Oxford, UK, 2011, pp. 151–160.