CHAPTER

1

Neurophilosophy or Philosophy of Neuroscience? What Neuroscience and Philosophy Can and Cannot Do for Each Other

M. Jungert

Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

INTRODUCTION

Ever since the rise of modern neuroscience in the 1980s, there has been controversial discussion about its potential influence on topics that have been traditionally seen as part of the domain of social sciences and humanities (see, e.g., Gold & Stoljar, 1999; Satel & Lilienfeld, 2013; Tallis, 2014).[a] The heated public and scientific debate on proposed neuroscientific solutions to the problem of the freedom of the will might be considered as the most prominent example (see, e.g., Mele, 2010, 2015; Schlosser, 2014; Walter, 2001). Moreover, the formation of a large number of neuro-hyphenated disciplines in the field of social sciences and the humanities such as neuro-theology,

[a] One result of this controversy is the formation of so-called "critical neuroscience" (see, e.g., Choudhury & Slaby, 2012; Slaby, 2010; Wolfe, 2014).

The Human Sciences after the Decade of the Brain.
DOI: http://dx.doi.org/10.1016/B978-0-12-804205-2.00001-X

neuro-psychoanalysis, neuro-education, or neuro-economics, to name just a few, shows the appeal and attraction of applying neuroscientific methods to traditional scientific fields.

In *philosophy*, two distinct ways of dealing with the problems and prospects of neuroscience have been developed in recent decades: On the one hand, the *philosophy of neuroscience* tries to apply methods and classical approaches from the philosophy of science to neuroscience, e.g., to shed light on its specific explanatory strategies. While this view is sometimes considered to be a more skeptical, critical, or even destructive one, so-called *neurophilosophy* takes a different approach. Here, neuroscientific findings are applied to classical philosophical issues such as the nature of emotions, the concept of morality, or the nature of consciousness, in order to develop empirically informed philosophical concepts and theories.

In this chapter, I am going to evaluate the premises and prospects of both approaches by discussing the following issues: I will start by reviewing the methods, theoretical assumptions, and explanatory aims of both the philosophy of neuroscience and neurophilosophy. In the next step, I will look into neurophilosophy's claim to integrate neuroscientific findings into philosophical theory by analyzing the relation between memory and personal identity. Based on this analysis, I aim to shed light on the more general question of what philosophy and neuroscience can and cannot do for each other.

WHAT IS THE PHILOSOPHY OF NEUROSCIENCE?

The so-called "philosophy of neuroscience" can be considered a branch of philosophy of science representing the ongoing trend to move from very general questions about science to more detailed discussions of particular issues of specialized disciplines. It applies classical concepts and questions from the general philosophy of science to the field of neuroscience. Research questions of the philosophy of neuroscience include: Is there a specific scientific method in neuroscience (Machamer, McLaughlin, & Grush, 2001)? Are there special kinds of explanations in neuroscience that differ from the types of explanations in other fields of science (Bechtel, 1994)? What is the impact of neuroscience on theories of human agency (Runyan, 2014)? Which concepts of causality or reduction are involved in neuroscientific explanation (Bickle, 2003)?[b]

[b] For a detailed general overview of the philosophy of neuroscience (see Bickle, Mandik, & Landreth, 2012).

One way of pursuing those questions is purely descriptive. If done that way, the agenda of the philosophy of neuroscience equals the approach of other specialized branches of the philosophy of science, e.g., like the philosophy of biology, physics, or psychology. In all those cases the main aim of philosophical investigation is to illuminate the field-specific ways of research and argumentation of an empirical discipline. Regarding neuroscience, one famous debate is the discussion about reductionism (see, e.g., Bickle, 2008; Craver, 2005; Schouten & Looren de Jong, 2007).

Generally speaking, the task of the philosophy of neuroscience is threefold:

First, it is the philosopher's job to discover and explicate the theoretical assumptions that are often more or less implicitly "woven into the fabric of empirical research" (Hyman, 1989, p. XIV). For example, Max Bennett and Peter Hacker state in their seminal work *Philosophical Foundations of Neuroscience*, "Many brain-neuroscientists have an implicit belief in reductionism. Few try to articulate what exactly they mean by this term of art" (Bennett & Hacker, 2003, p. 355).

Some philosophers of neuroscience see it as their task to make such implicit beliefs explicit in order to make them an object of investigation in the philosophy of science. This kind of explication work on background concepts can be considered a manifestation of the philosopher's general aim to dissolve conceptual puzzles and to confront others with their unquestioned or unconscious beliefs and assumptions. As in other fields of empirical research, neuroscientists mostly do not regard such basic questions as a matter of concern for themselves as they do not feel that these issues belong to their empirical core business. Therefore one main task of the philosophy of neuroscience in this context is to show that the proposed distinction between empirical core business and nonempirical sideline work is illusive as it ignores the fact that concepts, theoretical framework, and empirical investigation are intimately connected.

The second task is the distinction of different meanings of concepts that are either explicitly stated or implicitly used by neuroscientists. One example is the difference between ontological and explanatory reductionism in neuroscientific theories (Bennett & Hacker, 2003, pp. 355–366). Another one concerns the distinction between different meanings of "decision," which is one of the key terms in the debate about the freedom of human will (Walter, 2001, pp. 28–37). In those cases, the philosopher's job is to clarify the meaning of terms and the different ways of using concepts in order to make sure that discussions are really based on common concepts and not just circling around mock debates due to conceptual confusion. The heated free will debate between some neuroscientists and philosophers shows that many

misunderstandings and fruitless debates are due to conceptual confusion and could be avoided by clarification of concepts and by creating a common conceptual ground for fruitful interdisciplinary discussion.[c]

Finally, the third task of the philosophy of neuroscience is to discuss the plausibility of conclusions drawn from empirical data. For philosophers of science, one of the most irritating assumptions defended by some neuroscientists is the idea that far-reaching conclusions about human thinking and behavior can be more or less directly drawn from measurement results or brain imaging studies. Therefore it is the philosopher's task to analyze the structure of neuroscientific arguments and theories and to identify conclusions that are logically unsound or not supported by the presented data. Recent neuroscientific claims, among others made by Francis Crick, Gerald Edelman, or Antonio Damasio, offer plenty of examples (see Bennett & Hacker, 2003, pp. 68–74). One of them is the mereological fallacy that consists of ascribing mental states or complex abilities like deciding, believing, interpreting, perceiving, or thinking to the human brain as a part of a person instead of the person as a whole. In *The Astonishing Hypothesis*, Francis Crick gives a good example of this kind of fallacy: "What you see is not *really* there; it is what your brain *believes* is there [...]. Your brain makes the best interpretation it can according to its previous experience [...]. The brain combines the information provided by the many distinct features of the visual scene [...] and settles on the most plausible interpretation of all these various clues taken together" (Crick, 1995, p. 30). Philosophical analysis shows that this kind of ascription of psychological attributes to the brain simply does not make any sense. As Bennett and Hacker state, "The brain is not a logically appropriate subject for psychological predicates" (Bennett & Hacker, 2003, p. 72).

By categorizing such neuroscientific claims as confusing or even senseless, the philosopher is not just making a descriptive statement about neuroscience. In contrast to, for example, the reconstruction and description of theory formation in neuroscience, he takes a *normative* position toward his object of investigation. In a similarly normative way, he could try to show that certain correlations gained by neuroimaging studies do not reveal anything interesting about causal relations between brain states and behavior. The focus of the philosophy of neuroscience thereby switches from describing the structure of neuroscience to judging certain claims and eventually proposing alternative interpretations or models of explanation.

A survey of possible points of criticism concludes the characterization of the philosophy of neuroscience: Firstly, neuroscientists might complain that the philosophy of neuroscience represents exactly the

[c]See Kane (2011) for a broad overview.

kind of pointless "armchair philosophy" that tries to criticize empirical research from the outside without really knowing anything about its contents or methods.

Secondly, one could object that, while tackling foundational issues like explanatory strategies or concepts of representation, the actual research topics and empirical results are not at the center of attention. Instead of discussing current findings and helping to analyze, interpret, and consolidate the outcome of neuroscientific research, the philosophy of neuroscience only takes an interest in abstract conceptual and logical analysis. Moreover, it tends to lecture empirical scientists about issues that are remote from their core business or even completely irrelevant to their factual doing.

Thirdly, and finally, one could point at the one-sidedness of the philosophy of neuroscience. While it aims at analyzing and sometimes criticizing neuroscience, there is no attempt to consider neuroscience as a potential enrichment for philosophy, especially for the philosophy of mind. This ignorance, so the objection goes, inhibits productive interdisciplinary cooperation that is necessary for extensive research on the human mind and brain.

NEUROPHILOSOPHY: HOW TO COMBINE NEUROSCIENTIFIC FINDINGS WITH PHILOSOPHICAL THEORY

Against the background of this criticism, we can now turn to neurophilosophy as quite a different way of dealing with the prospects and challenges of neuroscience. The publication of Patricia Churchland's much debated book *Neurophilosophy* in 1986 can be seen as a major step in the development of this discipline. Churchland's approach is based on the assumption that close cooperation between neuroscience and philosophy and mutual integration of each other's findings and concepts are crucial for successfully studying brain and mind.

John Bickle, Peter Mandik, and Anthony Landreth describe Churchland's program as follows:

> She was introducing philosophy of science to neuroscientists and neuroscience to philosophers. Nothing could be more obvious, she insisted, than the relevance of empirical facts about how the brain works to concerns in the philosophy of mind. Her term for this interdisciplinary method was "co-evolution" [...]. This method seeks resources and ideas from anywhere on the theory hierarchy above or below the question at issue. Standing on the shoulders of philosophers like Quine and Sellars, Churchland insisted that specifying some point where neuroscience ends and philosophy of science begins is hopeless because the boundaries are poorly

> defined. Neurophilosophers would pick and choose resources from both disciplines as they saw fit. (Bickle et al., 2012)

In Churchland's understanding, neurophilosophy strongly differs from the philosophy of neuroscience. Neurophilosophers do not see themselves as critical observers of neuroscience, neither do they draw a sharp line between nonempirical matters of philosophy and empirical matters of science. In fact, they consider the philosophy of mind and neuroscience as closely related, intertwining, or even merging disciplines. However, the label "neurophilosophy" implies a unified concept while in fact it comprises several different approaches regarding methods and leading questions. The case of personal identity might help to illustrate this.

Taking a neurophilosophical approach to personal identity could, among others, be interpreted as follows:

- Firstly, one could attempt to transform philosophical criteria for personal identity into what Georg Northoff calls a "self-rating scale for empirical assessment of personal identity" (Northoff, 2004, p. 92). By doing this, abstract philosophical concepts are thought to be operationalized and converted into empirical concepts that can be applied to practical problems such as personality changes after brain surgery.
- Secondly, one could attempt to identify neural correlates of philosophical concepts, in this case personal identity. The potential results are sometimes thought to be a more precise replacement for allegedly cloudy philosophical concepts.
- Thirdly, one could attempt to integrate neuroscientific findings into a philosophical theory of personal identity in order to gain an empirically informed and enriched theory of personal identity. This approach is driven by the idea that—at least in some philosophical theories—there are elements of the theory that are open to the integration of empirical science.

I will now illustrate this third model by discussing the question of how the integration of memory research can be used to advance the philosophical debate on personal identity.

NEUROPHILOSOPHY IN ACTION: PERSONAL IDENTITY AND MEMORY RESEARCH

Memory is one of the most important features of human beings. Its importance becomes especially apparent in cases of severe amnesia or dementia, where loss of memory often seems equal to loss of identity

(Clark, 2010; Hoerl, 1999; Klein & Nichols, 2012). Although memory is a key element in many theories of personal identity from John Locke to Derek Parfit, the exact role that memory plays for the constitution and preservation of identity remains largely unclear (Schechtman, 1994). I claim that this negligence is, among others, due to diffuse concepts of memory and the lack of application of neuropsychological knowledge to philosophical theory (Jungert, 2013, 2015). I argue that both problems can be solved by integrating neuropsychological findings into philosophical theory, thereby creating an empirically informed new approach (Jungert, 2015).

As a start, it is helpful to consider the most prominent classification of memory systems, introduced by neuroscientist Larry Squire in the 1980s. Squire separates declarative memory from nondeclarative memory. While declarative memory is defined as "the capacity for conscious re-collection about facts and events" (Squire, 2004, p. 173), nondeclarative memory includes, among other things, skills, habits, and different forms of learning and is characterized "through performance rather than recollection" (Squire, 2004, p. 173). Although both systems are relevant to personal identity, I will limit the following discussion to declarative memory and especially to autobiographical memory as one of its subsystems. In contrast to semantic memory, the other subsystem of declarative memory, autobiographical memory is not memory of pure, neutral facts and knowledge. Instead, it contains significant emotionally charged memories about important events in the unique history of a person. Autobiographical memories can be remembered over a long period, often include a high level of detail, and enable the specifically human ability to mentally reexperience episodes from one's own past. By doing so, it becomes possible for human beings to evaluate and anticipate current and future actions based on experiences, personal preferences, and decisions from the past.

By combining the features and mechanisms of autobiographical memory discovered by neuroscience with the philosophical concept of biographical identity, it becomes possible to develop a much richer theory of the relationship between memory and identity compared to classical analytical approaches. I will give only one short example that concerns the emotional dimension of autobiographical memory. In many philosophical theories, memories are seen as countable units whose main purpose is to carry information over time (see Schechtman, 1994). They completely ignore the fact that memories are also a necessary precondition for experiencing one's life as a coherent, meaningful, and ongoing process, as they connect and organize remembered events in a way that allows for threading these different parts and that results in the ability of seeing one's life as a whole (Jungert, 2015, pp. 133–136).

Recent neuroscientific research elucidates this property of memory (Berntsen & Rubin, 2006). The emotional index attached to a certain memory at the time of encoding can become a part of the person's biographical identity. However, this emotional index might also be subject to change over time. The change can be caused by a new evaluation of the emotion and memory in question (Debus, 2007). If, for example, someone has changed his attitude toward smoking over the years, he might end up attaching his current aversion against smoking to his former memories, even though at the time of encoding he had been a passionate smoker and therefore attached positive emotions to these memories originally. Such cases demonstrate the reverse direction of influence from (current) self to (former) emotion and memories: "Not only is our sense of self based on memories of past experiences, [...] but our retrieval, recollection, and reconstruction of the past is, reciprocally, influenced by the self" (Schacter, Chiao, & Mitchell, 2003, p. 227). This modification of memories through the current self-image of a person usually happens without conscious awareness, resulting in implicit harmonization of remembered past and experienced presence.

Empirical findings such as these are extremely helpful for developing an advanced philosophical theory of the importance of memory for personal identity that is able to explain how persons are capable of developing narrative structures. It shows that these structures are necessary to understand one's own life as a more or less coherent story and demonstrates that memories can not only affect persons as carriers of information, but first and foremost as "transmitters of influence" (Wollheim, 1984, p. 101).

WHAT NEUROSCIENCE AND PHILOSOPHY CAN AND CANNOT DO FOR EACH OTHER

This example, although discussed very briefly, shows the enormous potential of this kind of neurophilosophy. The integration of neuroscientific findings into philosophical theory, done in the right way, can be of high value for both disciplines. For philosophy, this value consists of:

- Firstly, the chance to compare philosophical concepts to related concepts from neuroscience. In this context, "related" means that the concepts in question have some common content and are on a similar level of description. In the case of memory—as discussed—both disciplinary perspectives aim at a mental faculty and try to describe its importance for human beings. Therefore the comparison might show the philosopher that his own concepts are too narrow and fail to include important aspects of the object of investigation.

- Secondly, the potential revision of some elements of philosophical theories in reference to neuroscience. For instance, this holds for elements that claim to describe the actual functioning of certain cognitive powers. In the case of memory, such an element is the idea that memory can be considered as a warehouse that stores items of the past. Some philosophical theories of memory and personal identity rely on this idea implicitly or explicitly. Neuroscientific research shows that this idea is wrong in many respects. If the neuroscientific findings about the dynamics and inconstancy of memory are taken seriously by philosophers, this will also lead to new philosophical insights about memory and its role in the formation and structure of personal identity (Jungert, 2015).

For neuroscience, the value of neurophilosophical approaches consists of:

- Firstly, the discovery of implications of neuroscientific findings for topics that were originally outside the disciplinary focus. Examples include personal identity (Mathews, Bok, & Rabins, 2009), the nature of desire (Schroeder, 2004), or phenomenal consciousness (Clark, 1993).
- Secondly, the chance to make use of philosophical tools and methods of investigation. If neuroscientists are willing to engage in serious interdisciplinary dialog with philosophers, this provides the opportunity to benefit from conceptual and logical analysis. In contrast to some philosophers of neuroscience, neurophilosophers will consider the application of philosophical methods to neuroscientific findings not as a way of correction and falsification from a neutral outside perspective. In fact, they will see it as part of mutual learning and exchange that aims at a better understanding of complex mental phenomena by means of close collaboration on equal terms.

While it was my aim to mainly discuss the chances and positive effects of a certain way of understanding neurophilosophy, there are of course also certain problems and limitations. The outlined understanding could be characterized as "weak neurophilosophy," as it preserves the autonomy of the disciplines involved and abstains from strong claims regarding reduction or elimination. For some neuroscientists and neurophilosophers this conception will be way too careful and conservative.

In addition, the proposed way of integrating neuroscientific findings into philosophical theory might work well regarding topics like memory or perception. However, the intense debate on the possibility and meaning of "neuroethics" (see, e.g., Churchland, 2011; Farah, 2011;

Levy, 2009) suggests that it will not or only to a lesser extent work for other fields. To decide which fields can or cannot be an object of neurophilosophy and to determine boundaries and objections, one in turn needs help from the philosophy of neuroscience, and so the circle is complete.

References

Bechtel, W. P. (1994). Levels of description and explanation in cognitive science. *Minds and Machines*, *4*(1), 1–25.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Malden, MA: Blackwell.

Berntsen, D., & Rubin, D. C. (2006). Emotion and vantage point in autobiographical memory. *Cognition and Emotion*, *20*(8), 1193–1215.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Norwell, MA: Kluwer Academic Press.

Bickle, J. (2008). Real reduction in real neuroscience: Metascience, not philosophy of science (and certainly not metaphysics!). In J. Hohwy, & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 34–52). Oxford: Oxford University Press.

Bickle, J., Mandik, P., & Landreth, A. (2012). The philosophy of neuroscience. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Summer 2012 ed.). <http://plato.stanford.edu/archives/sum2012/entries/neuroscience/>.

Choudhury, S., & Slaby, J. (2012). *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*. Malden, MA: Wiley-Blackwell.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. S. (2011). *Braintrust: What neuroscience tells us about morality*. Princeton, NJ: Princeton University Press.

Clark, A. (1993). *Sensory qualities*. Cambridge: Cambridge University Press.

Clark, A. (2010). Memento's revenge: The extended mind extended. In R. Menary (Ed.), *The extended mind* (pp. 43–66). Cambridge, MA: MIT Press.

Craver, C. F. (2005). Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Science*, *36*(2), 373–395.

Crick, F. (1995). *The astonishing hypothesis: The scientific search for the soul*. New York: Touchstone.

Debus, D. (2007). Being emotional about the past: On the nature and role of past-directed emotions. *Noûs*, *41*(4), 758–779.

Farah, M. J. (2011). Neuroscience and neuroethics in the 21st century. In J. Illes, & B. J. Sahakian (Eds.), *Oxford handbook of neuroethics* (pp. 761–781). Oxford: Oxford University Press.

Gold, I., & Stoljar, D. (1999). A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences*, *22*(5), 809–830.

Hoerl, C. (1999). Memory, amnesia, and the past. *Mind and Language*, *14*(2), 227–251.

Hyman, J. (1989). *The imitation of nature*. Malden, MA: Blackwell.

Jungert, M. (2013). *Personen und ihre Vergangenheit. Gedächtnis, Erinnerung und personale Identität*. Boston, MA/Berlin: De Gruyter.

Jungert, M. (2015). Memory, personal identity, and memory modification. In R. Ranisch, M. Rockoff, & S. Schuol (Eds.), *Selbstgestaltung des Menschen durch Biotechniken* (pp. 129–140). Tübingen: Francke.

Kane, R. (Ed.). (2011). *The Oxford handbook of free will* (2nd ed.). Oxford: Oxford University Press.

Klein, S., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, *121*(483), 677–702.
Levy, N. (2009). Neuroethics: Ethics and the sciences of the mind. *Philosophy Compass*, *4*(1), 69–81.
Machamer, P. K., McLaughlin, P., & Grush, R. (Eds.). (2001). *Theory and method in the neurosciences*. Pittsburgh, PA: University of Pittsburgh Press.
Mathews, D. J. H., Bok, H., & Rabins, P. V. (Eds.). (2009). *Personal identity and fractured selves: Perspectives from philosophy, ethics, and neuroscience*. Baltimore, MD: Johns Hopkins University Press.
Mele, A. R. (2010). Testing free will. *Neuroethics*, *3*(2), 161–172.
Mele, A. R. (Ed.). (2015). *Surrounding free will: Philosophy, psychology, neuroscience*. Oxford: Oxford University Press.
Northoff, G. (2004). What is neurophilosophy? A methodological account. *Journal for General Philosophy of Science*, *35*(1), 91–127.
Runyan, D. (2014). *Human agency and neural causes: Philosophy of action and the neuroscience of voluntary agency*. Basingstoke: Palgrave Macmillan.
Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. New York: Basic Books.
Schacter, D. L., Chiao, J. Y., & Mitchell, J. P. (2003). The seven sins of memory: Implications for self. In J. E. LeDoux, J. Debiec, & H. Moss (Eds.), *The self: From soul to brain* (pp. 226–239). New York: Wiley.
Schechtman, M. (1994). The truth about memory. *Philosophical Psychology*, *7*, 3–18.
Schlosser, M. E. (2014). The neuroscientific study of free will: A diagnosis of the controversy. *Synthese*, *191*(2), 245–262.
Schouten, M. K., & Looren de Jong, H. (Eds.). (2007). *The matter of the mind: Philosophical essays on psychology, neuroscience, and reduction*. Malden, MA: Blackwell.
Schroeder, T. (2004). *Three faces of desire*. Oxford: Oxford University Press.
Slaby, J. (2010). Steps towards a critical neuroscience. *Phenomenology and the Cognitive Sciences*, *9*(3), 397–416.
Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(6), 171–177.
Tallis, R. (2014). *Aping mankind: Neuromania, darwinitis and the misrepresentation of humanity*. London: Routledge.
Walter, H. (2001). *Neurophilosophy of free will*. Cambridge, MA: MIT Press.
Wolfe, C. T. (Ed.). (2014). *Brain theory: Essays in critical neurophilosophy*. Basingstoke: Palgrave Macmillan.
Wollheim, R. (1984). *The thread of life*. Cambridge: Cambridge University Press.