

3:AM Magazine

- - [Home](#)
 - [Blog](#)
 - [Reviews](#)
 - [Essays](#)
 - [Interviews](#)
 - [Fiction](#)
 - [Poetry](#)
 - [Submissions](#)
 - [Twitter](#)
-

:: Article

Neuroethics

Interview by Richard Marshall.

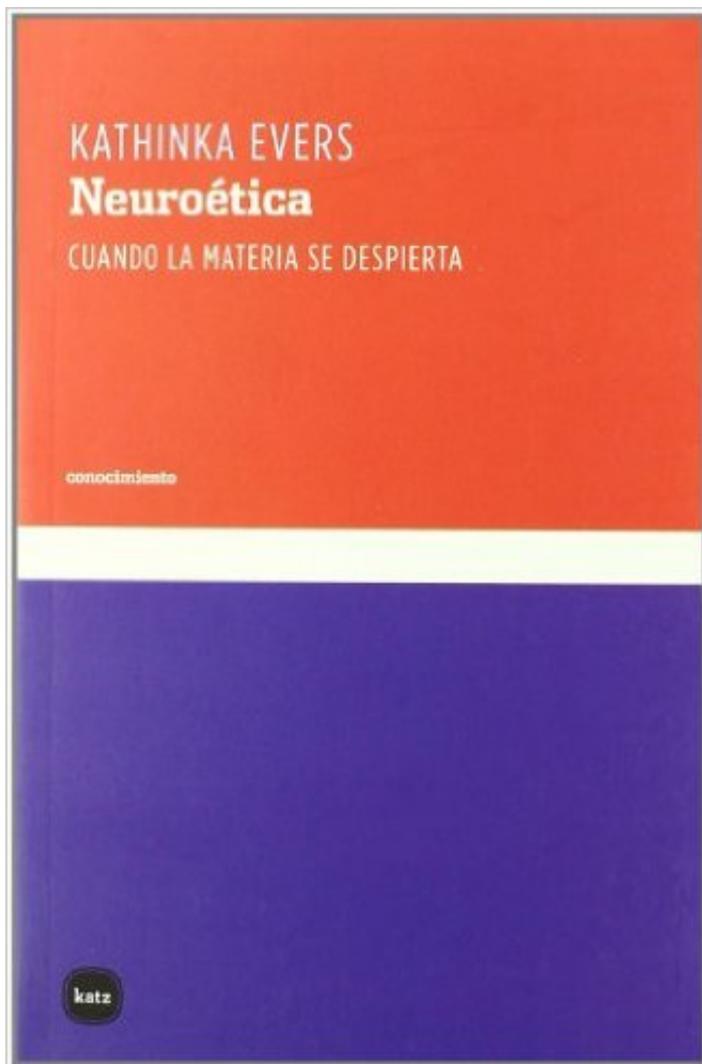


[Kathinka Evers](#) is a philosopher working at the cutting edge of [neuroethics](#). She thinks about what neuroethics is and what its questions are, about the distinction between fundamental and applied neuroethics, about the relationship between brain science and sociology, about how her approach avoids both dualism and naive reductionism, about mind-reading, about the ethical issues arising from disorders in consciousness, about brain simulation and its relation to philosophy, about whether tendencies in the brain lead to social or individualistic interpretation, about epigenesis, human enhancement, cognitive prosthetics and the singularity. Here's a post from the frontiers of neuroethics to take you through the xmas break...

3:AM: What made you become a philosopher?

Kathinka Evers: I was raised in a home where philosophy was a frequent topic for dinner conversations. Both my parents are academics, my father a philosopher, and he inspired an enthusiasm for philosophy in me at a very early age. Also for logic, which he taught me simultaneously with my learning to read and

write. Abstract thought appealed to me immensely ever since early childhood, and mathematics became the favourite topic at school from the first year and onwards. Later in life, in my early youth, I travelled extensively, and came into contact with profoundly different cultures, schools of thought, and values. This human diversity intrigued me in numerous ways, also from a philosophical perspective, and I started studying philosophy at the university. I began with logic and philosophy of science, took my doctoral degree in this domain (on the concept of indeterminacy in logic, philosophy and physics), but was also interested in moral and political philosophy. The latter was largely due to my upbringing that taught me how the social responsibility of an individual increases in proportion to her/his education, rather like a social debt: one must contribute to benefit society through one's education (it is not just a private playground created for one's own amusement). Philosophy of mind interested me deeply, but I was frustrated by the lack of empirical perspectives in the philosophical faculties when I was a student, where the road to hell was paved with empirical propositions! Yet it never seemed possible to me to understand the mind purely through a priori reasoning, ignoring the organ that does the job. On the other hand, brain science took scant interest in conceptual, philosophical analyses at the time, which seemed equally lopsided. Today the situation is fortunately different: philosophy and the neurosciences collaborate in a very fruitful manner. And that is why I now have turned my philosophical focus to studies of consciousness and [neuroethics](#).



3:AM: You're [working](#) in the field of [neuroethics](#). Can you sketch for the uninitiated what this is – this isn't just about translating ethics into brain science is it?

KE: It is partly that, but not only. Neuroethics is indeed concerned with the possible benefits and dangers of modern research on the brain. But neuroethics also deals with more fundamental issues, such as our consciousness and sense of self, and the values that this self develops: it is an interface between the empirical brain sciences, philosophy of mind, moral philosophy, ethics and the social sciences. It is the

study of the questions that arise when scientific findings about the brain are carried into philosophical analyses, medical practice, legal interpretations, health and social policy, and can, by virtue of its interdisciplinary character, be seen as a subdiscipline of, notably, neuroscience, philosophy or bioethics, depending on which perspective one wishes to emphasise. Such questions are not new, they were raised already during the French Enlightenment, notably by Diderot who stated in his *Eléments de Physiologie*: “C’est qu’il est bien difficile de faire de la bonne métaphysique et de la bonne morale sans être anatomiste, naturaliste, physiologiste et médecin...”. Moreover, ethical problems arising from advances in neuroscience have long been dealt with by ethical committees throughout the world, though not necessarily under the neuroethics label. Still, as an academic discipline labelled ‘neuroethics’, it is a very young discipline. The first “mapping conference” on neuroethics was held in 2002, and references to neuroethics in the literature were made little more than a decade earlier. These early articles described, for example, the role of the neurologist as a neuroethicist faced with patient care and end-of-life decisions, and philosophical perspectives on the brain and the self. Today, the pioneers of modern neuroethics have developed an entire body of literature and scholarship in the field of neuroethics that is rapidly expanding.



3:AM: In your view there are two types of neuroethics: fundamental and applied neuroethics and that the ‘fundamental’ aspect has been unrepresented in the field. Is your thought here that if the fundamental aspect isn’t worked out the applied aspect won’t be able to fully work?

KE: Yes. So far, researchers in neuroethics have focused mainly on the ethics of neuroscience, or applied neuroethics, such as ethical issues involved in neuroimaging techniques, cognitive enhancement, or neuropharmacology. Another important, though as yet less prevalent, scientific approach that I refer to as fundamental neuroethics questions how knowledge of the brain’s functional architecture and its evolution can deepen our understanding of personal identity, consciousness and intentionality, including the development of moral thought and judgment. Fundamental neuroethics should provide adequate theoretical foundations required in order properly to address problems of applications.

The initial question for fundamental neuroethics to answer is: how can natural science deepen our understanding of moral thought? Indeed, is the former at all relevant for the latter? One can see this as a sub-question of the question whether human consciousness can be understood in biological terms, moral thought being a subset of thought in general. That is certainly not a new query, but a version of the classical mind-body problem that has been discussed for millennia and in quite modern terms from the French Enlightenment and onwards. What is comparatively new is the realisation of the extent to which ancient philosophical problems emerge in the rapidly advancing neurosciences, such as whether or not the human species as such possesses a free will, what it means to have personal responsibility, to be a self, the relations between emotions and cognition, or between emotions and memory.

Observe that neuroscience does not merely suggest areas for interesting applications of ethical reasoning, or call for assistance in solving problems arising from scientific discoveries, as scientists of diverse

disciplines have long done, and been welcome to do. Neuroscience also purports to offer scientific explanations of important aspects of moral thought and judgment, which is more controversial in some quarters. However, whilst the understanding of ethics as a social phenomenon is primarily a matter of understanding cultural and social mechanisms, it is becoming increasingly apparent that knowledge of the brain is also relevant in the context. Progress in neuroscience; notably, on the dynamic functions of neural networks, can deepen our understanding of decision-making, choice, acquisition of character and temperament, and the development of moral dispositions.

3:AM: Some people in the social sciences express anxiety that brain science of this kind threatens the work of the social sciences. But you think it's a two way street, that the social sciences enrich neuroethics in certain areas. Is that right? Can you say something about this? Why shouldn't philosophers and social scientists be afraid of neuroscience?

KE: There are different possible reasons for this scepticism, which I can well understand, even though I regret very much when it leads to a rejection of collaboration across the fields. For, as you correctly point out, I consider the contribution of the social sciences and humanities deeply important, not only to neuroethics, but also to the natural sciences, notably neuroscience. Some of the reasons I see are the following four.

(1) Natural sciences have different degrees of explanatory power with respect to moral thought and judgment. The explanatory gap between our minds and our genetic structure is, I would say, larger than the explanatory gap between our minds and the architecture of our brains because the relationship between the latter two is closer than between the former in a manner that is explanatorily relevant. Simply phrased, neuroscience can explain more about why we think and feel the way we do than genetics does or can do. Even though an individual's genetic structure importantly determines who and what s/he becomes both physiologically and in terms of personality, genes only decide limited aspects of the individual's nature, and, at least so far as the mind is concerned, less than his or her brain structures do. In contrast, the brain is the organ of individuality: of intelligence, personality, behaviour, and conscience; characteristics that brain science increasingly is able to examine and explain in significant ways. Everything we do, think and feel is a function of the architecture of our brains; however, that fact is not yet quite integrated into our general world-views or self-conceptions.

The rapid neuroscientific advances may come to include profound changes in fundamental notions, such as human identity, self, integrity, personal responsibility, and freedom, but also, importantly, in neuroscience's models of the human brain and consciousness that has already moved away from modelling the brain as an artificial network, an input-output machine, to picturing it as awoken and dynamic matter. Through its strong explanatory power, neuroscience could be regarded as no less, and possibly even more controversial than genetics as a theoretical basis for ethical reasoning. Science can be, and has repeatedly been, ideologically hijacked, and the more dangerously so the stronger the science in question is. If, say, humans learn to design their own brain more potently than we already do by selecting what we believe to be brain-nourishing food and pursuing neuronally healthy life-styles, we could use that knowledge well – there is certainly room for improvements. On the other hand, the dream of the perfect human being has a sordid past providing ample cause for concern over such projects. Historic awareness is of utmost importance for neuroethics to assess suggested applications in a responsible and realistic manner.

(2) We know how genetics has lent itself to political prejudices of various kinds: conservative versus progressive, right-wing versus left-wing, male versus female, etc. Conservative ideologies trying to preserve the privileges of some specific class, race or gender sought support in genetic theories such as Mendelism, the theory of heredity emphasising the innate characteristics of the human being (some individuals could thus be said to be “born to poverty and servitude” and social reforms would make less sense). Progressive ideologies were inspired more by Jean-Baptiste de Lamarck's doctrine allowing for the inheritance of acquired characteristics and, by extension, of social flexibility. Attempts in the 1970s to establish socio-biology spurred intense controversies, and were attacked for joining the long line of

biological determinists. The reason for the survival of these recurrent determinist theories, argued critics, is that they consistently tend to provide a genetic justification of the status quo and of existing privileges for certain groups according to class, race or sex. That discussion became polarized in the extreme, where sociologists and biologists would sometimes reject all attempts to explain human identity and social life in any terms other than their own. Today, in contrast, biological and sociological explanations of human nature develop in parallel relations of complementarity rather than in stark opposition. Whilst some cases necessitate choices between the two perspectives (for example, if a specific disorder should primarily be medically or sociologically explained and treated), they are not seen as mutually conflicting generally. In some instances, of course, that peace may be frail. The ideological (and sometimes financial) interests in finding facts that suit a certain set of values are no less strong than they used to be, and their power to influence the scientific communities, through conditioned funding, political regulations, or by other methods has not diminished. Nevertheless, the all-out war of the trenches between biology and sociology appears to ebb away.

In contemporary neuroscience, the biological and socio-cultural perspectives dynamically interact in a symbiosis, which should reduce the tension further. This is particularly true of dynamic models of the brain arguing that whilst the genetic control over the brain's architecture is important, it is far from absolute; it develops in continuous interaction with the immediate physical and socio-cultural environments. The traditional opposition between sociology and biology is accordingly substituted by complementarity. An important task will be to unify different levels and types of knowledge combining technical and methodological approaches from distinct disciplines rather than to select one at the expense of another. Social sciences are extremely important for us to achieve an integrated and multi-level understanding of the brain. Note that homo sapiens is a species that spends large parts of its life developing the brain in response to learning and experience: culture leaves physical imprints on human brain architecture. And this symbiosis cannot be understood from a purely biological perspective.

(3) Another possible reason for scepticism has more to do with emotions or values. A central fear amongst those who reject the entry of natural science into moral philosophy and ethics is that the search for biological explanations of morality would somehow rob it of its moral, emotional, or human dimension, as people once feared the biochemical explanations of life. An equally central hope of those who see the development in a positive light is that the realisation that morality is a product of brains functioning in social, cultural environments, will empower, and enrich, the field of ethics. Surely, knowledge need not erode human dignity: if anything, the reverse ought to be the case. Even so, I can in view of recent history also well understand if the new socio-cultural-biological research of neuroethics sets ideological alarm clocks ringing. The solution, clearly, is to beware of any ideological misuse of theories developed and to maintain a high level of vigilance in this regard. Mistakes have been made in the past that should admittedly not be repeated. However, these mistakes have not only been of a political nature.

(4) Yet another obvious motivation for scepticism against neurobiological explanations of social phenomena, such as moral thought and judgment, is the harsh destiny that awaited the concept of the conscious mind when science secularised that area of research and placed the human mind firmly in nature. Schools of thought emerged that did indeed rob the conscious mind of both meaning and content, scientifically speaking. In its eagerness to escape dualism, science in the 20th century became to no small extent psychophobic and that is important to bear in mind when we discuss the relevance and value of neurobiological explanations of thought and judgment.

The sciences of mind suffered from severe psychophobia until late in the 20th century, and it is perfectly legitimate not to want neuroethics to cross the same desert. The doctrines of behaviourism invaded psychology and were followed by naïve eliminativism and naïve cognitivism. One eliminated the mind; the other emotions and the brain from their pursuits, and the result was, of course, seriously lop-sided. I consider an interesting study of psychology the question why any thinking being would want to reduce its own mind to a behaviouristic slot-machine, or indeed to any machine, organic or otherwise. And, as Joseph LeDoux asks: Why would anyone want to conceive of minds without emotions?

The scientific situation today has evolved considerably from what it was a century, half a century or even a decade ago. Mind science is far less psychophobic (if at all), and radical eliminativism with respect to consciousness has lost most of the ground it once possessed. Modern neuroscience is in important ways and measures non-eliminativist both ontologically and epistemologically: it neither denies the existence of mind (conscious or non-conscious), nor does it deny that the mind is an important and relevant object of scientific study, nor does it necessarily presume to explain subjective experience without the use of self-reflection. The image of the brain that some contemporary neuroscientists offer is as far from behaviourism or the mind-machine model in which the brain's activity is depicted in an input-output manner as it is from the religious notions of an immaterial soul. Psychophil science has taken the ground.

Scientific theories about human nature and mind in the 19th and 20th centuries were occasionally caught in two major traps: ideological hijacking, and psychophobia in the form of naïve eliminativism and naïve cognitivism. In order to avoid repeating these mistakes, neuroethics needs to build on the sound scientific and philosophical foundations of informed materialism. This is a concept originally coined in chemistry (by Gaston Bachelard) that has been extended to neuroscience (by Jean-Pierre Changeux) and to philosophy (by KE) in a model of the brain/mind that opposes both dualism and naïve reductionism. This model is based on the notion that all the elementary cellular processes of brain networks are grounded on physico-chemical mechanisms and adopts an evolutionary view of consciousness as a biological function of neuronal activities, but describes the brain as an autonomously active, projective and variable system in which emotions and values are incorporated as necessary constraints. Due to the way in which our capacity-limited brains acquire knowledge of the world and ourselves, informed materialism acknowledges that adequate understanding of our subjective experience must take both self-reflective information and data gathered from physiological observations and physical measurement into account. Informed materialism depicts the brain as a plastic, projective and narrative organ evolved in socio-biological symbiosis, and posits cerebral emotion as the evolutionary hallmark of consciousness. Emotions made matter awaken and enabled it to develop a dynamic, flexible and open mind. The capacity for emotionally motivated evaluative selections are what distinguish the conscious organism from the automatically functioning machine. And herein lies the seed of morality.



3:AM: How do you see the relationship between the empirical research and the philosophical analysis of concepts of concepts such as ‘consciousness’? Presumably the analysis impacts on the research? I guess this is really a question about what role you think philosophy has in neuro science?

KE: If I may begin by paraphrasing Immanuel Kant: Conceptual analysis of mind without empirical content is empty; empirical analysis of mind without conceptual analyses is blind. The basic role of philosophy (as I see it) is to clarify concepts, theories and arguments; reveal underlying assumptions of suggested theories and data, as well as their implications both theoretically (e.g., epistemologically) and practically (e.g., ethically and socially). It is a help to interpret correctly the results of empirical experiments, such as what fMRI scans actually “reveal” in the brain, or what it means to say that we can “communicate” neurotechnologically with patients in vegetative states, or when we “read minds” without overt behaviour or speech. Philosophy is in quest of meaning, bringing understanding of concepts to a higher level, developing theories that are more refined, clearer, and more coherent. Without philosophy, neuroscience stands a much greater risk of misinterpretations and other errors.

3:AM: Philosophers like [Goldman](#) and [Carruthers](#) think about mind reading: from the perspective of neurophilosophy what do you think are the possibilities and limits of this?

KE: The possibilities of neurotechnological mind-reading that we have today allow access to mental states without 1st person overt external behaviour or speech. With the advancement of decoders of cerebral activity it is very likely that in the near future we will see a rapid progression in the capacity to observe – without mediation of language – contents of the others' mind. We are seemingly able to efficiently use a subject's cerebral cortex for rapid object recognition, even when the subject is not aware of having seen the recognized object. This may be extended as a great promise to the domain of dreams, to observe in real time the content of a visual narrative during sleep. We might be able to infer a myriad of simultaneous intentions whose deliberation process to reach explicit agency is not tangible even to the same subject. We might be able to use this technology in medical situations (most notably in patients with consciousness disorders) where this might be the only available tool to infer another person's will. Certainly, applications in commercial setups to control objects (games, cars, airplanes) that are currently under massive development will become more frequent and effective.

There is a logical limit to these pursuits, in that an individual cannot wholly share another's experience without merging with it. Their distinction necessarily introduces a filter, an interpretation that individuates their respective points of view. In other words, by virtue of our distinction we have a private room that cannot logically be violated. The presence of this logical limit says nothing about the extension of our privacy, except that it isn't null. It does not exclude that our unalienable privacy may be extremely small. Moreover, it does not entail that we need have privileged access to our own experiences: the fact that there is an essential incompleteness in any other person's knowledge or experience of you does not mean that there is no, or less, incompleteness in your own self-understanding. To the contrary, it is possible that a brain decoder may access more information about, say, the intention of a subject than that which may be simply accessed by introspection.

The specific benefits of neurotechnological mind reading include the following:

- For a person who suffers from behavioural incapacity for communication, the prospect of neurotechnological mind reading opens up promising vistas of developing alternative methods of communication.
- The development of these techniques holds promises of important medical breakthroughs, notably improvements in the care and therapeutic interventions of patients with disorders of consciousness.
- For those – parents, paediatrics, and others – interested in understanding the infant pre-verbal mind, the research opens promising vistas.
- For radiology or satellite reconnaissance, notably, optimizing image throughput by coupling human vision with computer speed is a promising area of research.
- For philosophy of mind and all sciences of mind, whether they are clinically orientated or not, the research into neurotechnological mind reading is exciting and appears theoretically promising.

The development of mind reading can also be perilous, however, increasingly so if or when the techniques advance. There is, notably, a risk for misuse as a consequence of hypes, exaggerations, or misinterpretations, and a potential threat to privacy unknown in history. At present, the possibilities of neurotechnological mind reading are so rudimentary that the techniques pose threats to privacy mainly in the form of misuse, but this threat might expand and increase if the techniques are refined. In that context, the question arises: who is best placed to know what goes on in a person's mind? Who is authorized to say? Does the 1st person have privileged access, or the one to perform/interpret the cerebral measurements? Already, a person's unconscious recognition of an image can be detected. How far can that be taken? Today, at the present level of science and technology: not far. Yet in the future, if better models and measurements of brain functions and mental contents are developed, the day could come when another, with the use of neurotechnology, enters your mind further than you can yourself. Is that a threat, or a promise? How we evaluate the integrity of our mind depends in part on our trust in others and our views on society: in which society we live; and which society we want to see develop in the future.

3:AM: You've examined the ethics of treating people who have disorders in consciousness. Can you describe some of the conditions you are discussing and say what ethical issues arise from these situations?

KE: Three of the main diagnoses of disorders of consciousness (DOCs) are Minimally Conscious State (MCS), Vegetative State (VS), and Coma. Their distinction is often described in terms of two dimensions: wakefulness (referring to arousal and the level of consciousness) and awareness (referring to the content of consciousness and subjective first-person experience). Patients who are in MCS can, as the name suggests, show some signs of awareness: some MCS patients may retain widely distributed cortical systems with potential for cognitive and sensory function despite their inability to follow simple instructions or communicate reliably. In contrast, the diagnostic criteria of coma exclude the presence of awareness and responsiveness as well as wakefulness. Coma is defined as a state of unarousable unconsciousness due to dysfunction of the brain's ascending reticular activating system (ARAS), which is responsible for arousal and the maintenance of wakefulness.

The diagnostic criteria of VS likewise exclude the presence of awareness; however, these patients can move, open their eyes, or change facial expressions. By virtue of these bodily states and movements, VS is considered to be one of the most ethically troublesome conditions in modern medicine, since bodily states can be taken to be indexes of mental states, something that may cause psychological problems for the next of kin, and diagnostic doubts in the caregiver. Recent studies of DOC patients prompt a question that has ethical implications: is it accurate to describe patients with VS or coma as totally unaware of themselves and their environment? Or do some of those patients possess preserved mental abilities undetected by standard clinical methods that exclusively rely on behavioural indexes? Numerous ethical issues arise in this clinical context, notably: the problem of misdiagnosis, assessment of detected residual consciousness in DOC patients and (if applicable) the interpretation of their 1st person experiences, developing communication with these patients (if possible), decisions on adequate treatment, adapting the living conditions of these patients taking their possibilities of enjoyment or suffering into account and providing support for those who are close to the patient, and the question whether life-sustaining care should be discontinued in case the patient suffers.

3:AM: Do the ethical and legal concerns overlap in these patients?

KE: In some cases, yes, for example the concern whether to discontinue life-sustaining care if a patient is believed to suffer. But all ethical concerns are not legally regulated. And all legal regulations are not as such ethical.

3:AM: A technological fix is an obvious thing to want if you're an engineer or scientist. So brain simulation seems equally an obvious thing to try if we're trying to fix problems of consciousness. But simulation raises interesting philosophical questions in you doesn't it. So first could you sketch out what simulation in this context looks like?

KE: To my knowledge, simulation is not yet used in the studies of consciousness disorders, but this could be an interesting future development. I am not an expert on simulation. I only began studying it a couple of years ago when I became involved in the Human Brain Project. What I say below are ideas published and co-authored with a colleague in neuroscience, Yadin Dudai. I will begin by discussing the goals of simulation. In experimental science, simulation is one of the four meta-methods that subserve systematic experimental research. These are: observation, the most fundamental of all the experimental methods, clearly preceding modern science; intervention, currently the most popular method in reductive research programs, with the aim of inferring function from the dysfunction or hyperfunction of the system; correlation of sets of observations or variables extracted from the observations, or of the effect of interventions, in order to identify links between explicit or implicit phenomena and processes; and simulation, to verify assumptions, test heuristic models, predict missing data, properties and performance, and generate new hypotheses and models in which these experimental meta-methods are commonly enwrapped (the order in which the meta-methods are listed above does not of course imply that they are used in that order in realistic research programs). Simulation is hence used here to provide a proof of

concept in the course of research and to promote and achieve understanding of the system.

When scientists use simulation in this manner, they either explicitly or implicitly assume that in order genuinely to understand a system, one should be able to reconstruct it in detail from its components. This assumption resonates with a maxim of scholastic philosophy, resurging in Vico (1710): only the one who makes something can fully understand it. 'Understanding' as a cognitive accomplishment is intuitively understood but its meaning(s) in science is debated. For many scientists, understanding refers to the ability to generate a specific mental model (or a more encompassing theory) that permits predictions based on scientific reasoning concerning the behavior of the system under different conditions at the specified or additional level(s) of description. One particular point that is highly pertinent to a philosophical discussion of simulation is the level of epistemic transparency assumed to be required to reach understanding of the system. In other words, what is the magnitude of the epistemic lacunae or 'gaps in understanding' that one is willing to tolerate in a simulated model while still claiming that the simulation increases scientific understanding at the pertinent level of description. This point is particularly relevant to the understanding of complex, nonlinear systems such as the brain, i.e., systems with emergent properties in which the behavior of the system is unaccountable for by the linear contributions of the components.

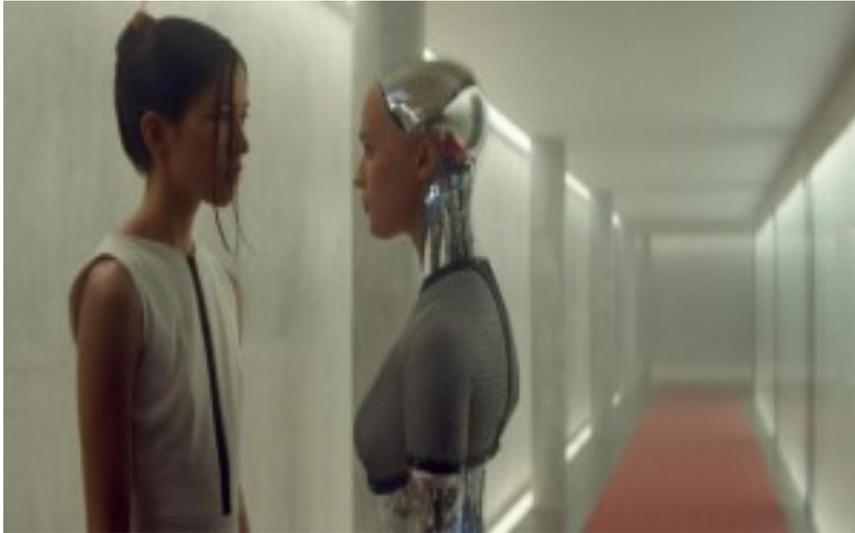
In the brain sciences, understanding is currently realistic with respect to only a limited number of basic neural operations and brain functions. Some types of simulations, however, have a long history of being a productive tool in testing and advancing partial understanding of the mechanism of action of neural systems. They are also considered in attempts to impact the development of artificial computational systems and brain inspired technologies.

For instance, since the outset of the powerful reductionist approach to the neurobiology of plasticity and memory, perceptual input and motor output of neural systems have been simulated by substitution with direct electrical stimulation of nerve fibers and of identified sensory or motor nerve cells, respectively. In this type of approach, the artificial agent that simulates or functionally substitutes the natural component is further used to manipulate the system in order to demonstrate that the modeled state or process are indeed functioning as expected. Hence the input of the conditioned stimulus (CS) in Pavlovian or instrumental conditioning is replaced with artificial stimulation of the natural input to prove that identified parts of the neural circuit in vivo fulfill or at least take part in the role assigned to them in a model of the functional nervous system.

Another philosophically important question concerns the nature of the object: what is the 'brain' that brain simulation targets? In real life, brains do not live in isolation. In other words, brains are complex adaptive systems nested in larger complex adaptive systems. They reside in bodies. The interaction between the brain and the other bodily systems is, in reality, impossible to disentangle. Our brain gets and sends information to all other bodily systems, and its state at any given point in time is determined to a substantial degree by this interaction. That the brain is a brain-in-a-body cannot be ignored in considering the goal to simulate the realistic brain. But the brain-in-a-body at any given point in time is in fact the outcome of the individual experience accumulated over the period preceding this specific point in time. In simulating the brain, one has therefore to consider the experienced-brain-in-a-body. Neglecting experience sets a severe limit on the outcome of brain simulation. On the other hand, taking experience into account necessitates simulating real-life contexts, a daunting task per se, specifically given that part of the real-life experience is the interaction over time with the functioning body. In specifically discussing a hypothetical human brain simulation, it seems logical to limit the goal to the individual, yet without ignoring the relevance of the natural, social and cultural interactions and contexts over time. Therefore, the question how this limitation may affect the adequacy of large-scale simulation attempts in due time and their results must be borne in mind. Some key considerations are the following:

Scarcity of knowledge: Collection of data for realistic large-scale brain simulation is not trivial. Even a highly productive large experimental laboratory investigating the mammalian brain can produce only limited amounts of data. Federating data from different labs has to take into account that even small

differences in methodology and conditions can mean a lot in terms of neuronal state and activity, and different labs seldom if ever use exactly the same conditions and protocols. The invariants identified under these conditions may mask important features. This complicates the ability to merge data from different sources without losing important information. Heterogeneous data formats also present an obstacle in sharing. As far as data required for human brain simulation are concerned, it is sufficient to note that cellular physiology data are scarce and obtainable from patients only. Functional neuroimaging using fMRI has limited spatiotemporal resolution which currently constraints its applicability to high-resolution brain simulation, though is useful in obtaining important information on the role of identified brain areas and their functional connectivity in perceptual and cognitive processes. One possibility to bridge the gap from the cellular to the cognitive is to use data from the primate brain, but these data are also yet insufficient for the purpose of large-scale brain simulation.



Epistemic opacity: Is the aforementioned Vico maxim, that posits that one can only understand what one is able to build, i.e. that truth is realized through creation, applicable to computer simulation of complex systems? Having fed the information and let the machine run the computations involving strings of equations and come up with emergent properties, do we really understand the system better as long as part of the process is epistemically opaque? And what is it that creates the opaqueness, given that we in fact wrote the equations – the numerical iterations, high dimensionality, nonlinearity, emergence, all combined? This brings us back to the meaning of ‘understanding’. Some will note that even in daily life, we claim to understand natural phenomena without really mentally grasping their inner working. For example, we predict that if we release a ball from a tower, the ball will fall because of gravity. But is the attraction of physical bodies transparent to us epistemically, or is our sense of understanding due to habituation with the phenomenon or the physical law? As noted above, the acceptable magnitude of epistemic opacity in a computer simulation that can predict the outcome of the behavior of the system, is for the individual scientist to decide, and will probably vary with the professional training and the level of description and analysis.

Computing power: The computing power required for large-scale simulation of a mammalian brain is yet unavailable. Exascale-level machines are required, that, if pursued by current technology, will demand daunting amounts of energy. However, given the fast pace of advances in computer technology, this issue will probably resolve prior to the resolution of the scarcity of knowledge problem mentioned above.

The toll of data sampling: Attempts at large-scale brain simulation differ with regard to their reliance on realistic and detailed brain data, but all currently rely on limited sampling and statistical typification. It is one thing to sample phenomena in experiments in search for mechanisms and to classify the data to facilitate understanding, another to rely on the sampling to faithfully build the system anew. The possibility cannot be excluded, hence, that important properties of real-life neurons in vivo are concealed or minimized in the process. It is noteworthy that relying on extracted invariants may result not only in missing data but also in going beyond the data, because of potentially erroneous generalizations. It is also

of note that such methods may reduce the ability to rely on the simulation to perform new fine-grained experiments in silico ('higher order simulation'), which is contemplated as one of the contributions of brain simulation (i.e. replace in vivo or in vitro experiments that are complex, time consuming and cause animal suffering). Further, it may result in a situation in which the outcome of an in silico experiment will have to be verified in vivo after all.

Reality checks: Large-scale simulations are expected to involve iterations in which the performance of the simulated systems is evaluated by benchmarks. However, scarcity of knowledge may raise doubts concerning the suitability of such benchmarks, as we do not yet know in most cases whether the correlation sought by us of the activity of an identified circuit with specific physiological or behavioral performance indeed reflects the native function of the circuit. For example, are place cells primarily sensitive to spatial coordinates, or amygdala circuits to fearful stimuli? Lack of knowledge on the native computational goal may result in optimizing simulations to misguided or secondary performance. On the other hand, one may consider using the fit of simulations to selected benchmarks to explore computational goals of the native circuit.

Representational parsimony: Much of our scientific progress, understanding and intellectual joy stems from our cognitive ability to extract and generalize laws of nature. Describing the universe in a minimal number of equations is often equated not only with ultimate understanding but also with beauty. If we aim to reproduce details in simulations, do we still advance in 'understanding' in that respect, or just imitate nature? Proponents of large-scale simulations will claim that the reproductions of the details is practiced in order to extract new laws that may emerge from the simulation. Besides raising again the issue of epistemic opacity, a more practical question comes up: Should we expect a small set of laws to describe a complex adaptive system like the brain? Some will say that this depends on the level of description. The brain can be considered as a community of organs with different functions and phylogenetic history, which renders the hope to understand in detail the operation of each by the same task-relevant computations doubtful. It still leaves open the possibility that some basic principles of brain operation are explainable by a unified theory. But this depends on the level of description. One may claim that we already understand some fundamental principles of brain operation, for example, that spikes encode and transmitters convey information, but this level of description is obviously not what brain scientists have in mind in trying to 'understand' the brain. It is of note that high parsimony in realistic models has the potential to ameliorate epistemic opacity.

3:AM: How do you think simulations and philosophy should be integrated in this approach? What should we be trying to achieve?

KE: For example: Science and society should aim to benefit from contemplating the future and prepare for it, even if this future is not necessarily around the corner. Suppose, for the sake of argument, that the brain and computer sciences combined will indeed be able one of these days to come up with a simulated human brain. What questions will we face?

Similarity of the simulation to the original: If the simulation is in silico, there is the obvious dissimilarity that the simulation versus the original are two different substrates. The relevance of this dissimilarity can be expected to vary with theoretical frameworks and contexts. If, for example, one takes the hypothetical position that consciousness can only arise in a biological organism (see below), the relevance of the difference in substrate will be very high, since it will entail the further dissimilarity of being capable versus incapable of possessing mental states.

The issue of similarity can also be raised, however, within an in silico universe. Suppose, for the sake of argument, that we succeed in some imaginary future to generate a faithful simulation of the native human brain that is embodied in neuromorphic devices, embedded, for example, in humanoid robots. Will we be able to create legions of identical brains? The question of similarity of such artificial copies of the human brain can be dissected in terms of internal structure, or spatiotemporal location. The question can be broken up into two levels: type similarity, i.e. will the process generate a type of machine that is similar to

a generic brain, and token similarity, i.e. will the process generate specific copies of an individual brain. In that case, in theory, type similarity is a possibility. Yet token similarity is a different question. That issue can benefit from the classic discourse in analytic philosophy, related to Leibniz's principle (or 'Law') of The Identity of Indiscernibles. This principle states that if, for every property F, object x has F if and only if object y has F, then x is identical to y. In other words, no two distinct things exactly resemble each other, because if they share all intrinsic and all relational qualities (e.g. spatiotemporal coordinates) they would then be not two but one. They can, however, share all intrinsic qualities and yet be relationally, e.g. spatially or temporally, distinct. Formally we do not expect, therefore, even a future perfect brain simulation project to produce token identity.

3:AM: Will consciousness emerge? When mental states of the human brain are considered, consciousness commonly comes up in the discussion. Can consciousness be simulated?

KE: A dominant conceptual framework posits that mental states are brain states. Will (or must) intrinsically identical brains have identical mental states? Will distinct simulated brains with identical mental states be considered distinct 'individuals'? Will they be able to read each other's 'mind'? (Presumably, yes, if they know their intrinsic identity and the answer to the first question is affirmative.) Will they significantly differentiate even if they share identical experiences? Many brain scientists will posit that they will diverge over time because they consider the possibility that at least some systems in the brain will be of the type that is sensitive to minuscule deviations in the initial states (this also reflects on the improbability of token identity, see above).

Further, mental states may not correspond on a one-to-one basis to brain states; or mental states are functions of the brain with some other relation to brain states, for example, they are only supervenient or consequential to brain states, come along with them, but are not necessarily entailed by them in a one-to-one relation, in a way that brain research can not yet account for. But could the computer be conscious at all? At present, available evidence justifies only a rather tame hypothetical stance: If consciousness is necessarily an outcome of a certain type of organization or function of biological matter, then brain simulation will never gain consciousness; whereas if consciousness is a matter of organisation alone, e.g. extensive functional interconnectivity in a complex system, then it might arise in simulations in silico.



3:AM: How would we recognise whether a future brain simulation is conscious or not?

KE: Two main types of approaches can be raised. The first, a Turing-type test for a conscious entity. Yet by itself this is insufficient, because we can easily imagine a computer being able to mimic the expected responses of a conscious entity without experiencing consciousness. The second, provided we assume faithful imitation of the relevant native brain activity, identify activity signatures that reflect conscious awareness in the human brain. This is in principle similar to the way one attempts to identify sleep and dreams objectively, by looking for characteristic brain activity signatures. But on the one hand, we do not yet know such signatures; on the other, even if they are identified, they may not exhaust signatures of conscious awareness in a simulated system. A pragmatic heuristic approach could be combination of two elements, still short of a sufficient condition. One, a Turing-type test; the second, activity signature in the simulated entity that fits the one expected in the original biological brain, and is time locked to the responses taken to reflect conscious behaviour.

3:AM: Is realistic human brain simulation possible in the absence of consciousness?

KE: It is possible to consider brain simulation without the question of consciousness arising. However, when processes in the brain are simulated that are conscious in the human being (for example, declarative emotion), the question arises: if consciousness is not simulated, how adequate can that simulation be?

To illustrate, one of the proposed goals of human brain simulation is to increase our understanding of mental illnesses, and to ultimately simulate them in theory and possibly in silico, the aim being to understand them better and to develop improved therapies, in due course. But how adequate, or informative, can a simulation of, say, depression or anxiety be, if there is no conscious experience in the simulation? The role of consciousness and the effects of this role on the outcome of simulation of human brain faculties will be important to assess in this context.

3:AM: So: what can we gain from discussing brain simulation?

KE: Although the road to simulation of human brain, or even only part of its cognitive functions, is long and uncertain, on this road much will be learned about the mammalian brain in general and about the feasibility of transformation of some efforts in the brain sciences into big science. New methodologies and techniques are expected as well that will benefit neuroscience at large and probably other scientific disciplines as well.

But given the expected remoteness of the ultimate goal, why should we engage in discussing some of its conceptual and philosophical underpinnings now? Big science brain projects provide an opportunity to assess and preempt problems that may one day become acute. In other words, we can use the current attempt to simulate the mammalian brain as an opportunity to simulate what will happen if the human brain is ever simulated.

It is rather straightforward to imagine the types of problems a simulated human brain will incite, should it ever become reality in future generations. They will range from the personal (e.g. implications concerning alterations of the sense of personhood, human identity, or anxiety and fear in response to the too-similar other); social (e.g. how shall the new things be treated in terms of social status and involvement, the law, or medical care); and ethical (e.g. if we terminate the simulated brain, do we 'kill' it, in a potentially morally relevant manner?). These problems also require foresight of safety measures to ensure that in due time, the outcome of ambitious brain projects do not harm individuals and societies. But most of all, by discussing the potential implications of such projects now, we contribute to the sense that scientists as individuals and science as a culture should take responsibility for the potential long-term implications of their daring projects.

3:AM: You think there are tendencies in the brain that place us in a predicament of whether we go social or whether we go individualistic. Could you first sketch out for us what it is about the brain causes the predicament and why this is significant?

KE: I think we are fundamentally social as well as individualistic. The problem, simply phrased, is that we may be biologically unable to apply certain values that we intellectually endorse, because we are

imprisoned in a smaller context. Let me try to explain this more fully.

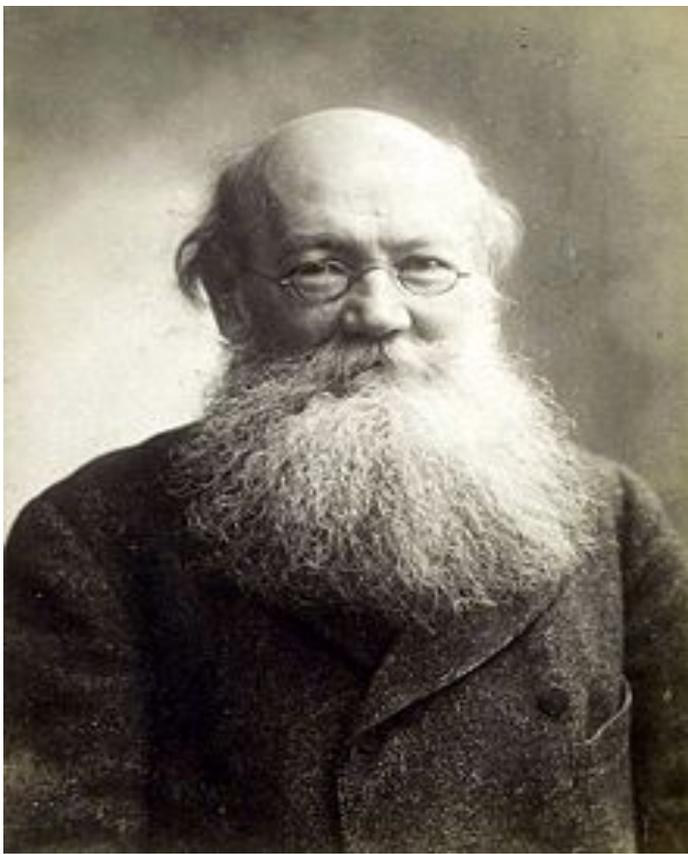
Self-awareness can only develop through social interaction. The human brain is fundamentally social, and develops in natural and social contexts that strongly influence its own architecture. In social creatures, self-interest is a source of interest in others, primarily those to whom the self can relate and with whom it identifies, such as the next of kin, the clan, the community, etc. In intelligent social species such as the human, the “I” is extended to endorse the group, “we”, and distinctions drawn between “us” and “them”. Sympathy and aid is typically extended to others in proportion to their closeness to us in terms of biology (e.g., face recognition, or racial outgroup versus ingroup distinctions), culture, ideology, etc.

Evolution seems to have predisposed social animals to develop norms and rules for their behaviour, for example assistance within the group, where failure to follow social rules or conventions can have serious consequences. However, even in favourable conditions, we are not necessarily biologically capable of following all social rules. Ample evidence shows how brain dysfunctions or damages can underlie a multitude of cognitive, emotional and behavioural disabilities, including self-indifference and social or moral incapacity, and how the structure of the supposedly healthy brain may also render some norms more or less inapplicable in practice.

Our capacity for understanding others or for sympathising with them is dependent on brain functions. Compassion, for example, requires an intellectual capacity to understand the other, as well as an emotional capacity to care about the other. Both of these functions in the brain can be disordered or damaged, and even in brains that are supposedly neither, these functions are pronouncedly selective.

The neurobiology of empathy, here understood as the ability to apprehend the mental states of other people, is today subject to extensive research suggesting that this ability is a complex higher cognitive function with large individual and contextual variations that depend on both biological and socio-cultural factors. In some individuals, the capacity for empathy is seriously reduced. Those who suffer from Asperger’s Disorder, for example, are largely unable to understand other people’s minds, to envisage how they think or feel. Still, to the extent that they succeed, they are able to sympathise. Individuals with a psychopathy disorder find themselves in the reverse situation: the structure of their brains makes them less able to experience certain emotions, such as sympathy, guilt, shame, or other morally relevant emotions, but they can nevertheless be well able to envisage what other people feel.

There is, accordingly, a biological distinction between moral and social understanding (knowing what is considered ‘right’, ‘wrong’, ‘good’, ‘bad’, etc.) and moral or social emotion, such as sympathy, embarrassment, shame, guilt, pride, etc.



Pjotr Kropotkin, whose idealistic interpretation of history made him see voluntary mutual helpfulness and sympathy in his studies of nature, emphasised in sharp contrast to Thomas Huxley the positive aspects of nature: the tendency to altruism and mutual aid that stems from our natural capacity for sympathy with others. However, Kropotkin's and Huxley's images can be wedded; for when sympathy and mutual aid is extended within a group, they are also (de facto) withheld from those that do not belong to this group. In other words, interest in others is ordinarily expressed positively or negatively through either sympathy or antipathy directed to specific groups – but very rarely, if ever, are attitudes extended to universal coverage, for example as attitudes towards the entire human species, let alone towards all sentient beings. Our standards for normality versus mental illness or disorder reflect this feature.

Emotional inabilities can be diagnosed as signs of a psychiatric disorder, but there is no corresponding diagnosis of a person who is indisposed to feeling shame or sympathy in relation larger groups, e.g. humanity, so long as that person remains capable of relating “normally” to individuals. This is a rule rather than an exception, if we look at the standard works defining mental disorder, such as the DSM IV. In other words, our diagnostic criteria for mental disorders reflect relationships between individuals rather than between an individual and a large group. This may be realistic, but also reflects a serious human predicament.

Even in human beings that are not diagnosable as suffering from a brain disorder or mental illness, understanding does not entail compassion but is frequently combined with emotional dissociation from “the other”. We can easily understand, say, that a child in a distant country probably reacts to hunger or pain in a way that is similar to that in which our own country's children react to it, but that does not mean that we care about the children in equal or even comparable measures. Indeed, if understanding had entailed sympathy, the world would be a far more pleasant dwelling place for many of its inhabitants.

Humans are biologically natural sympathisers with the groups to which they belong, and can understand groups to which they do not belong, but they are not equally disposed to sympathise with them. To the contrary, we behave towards the greater part of the world in a manner that may have suggested a psychopathic disorder had it been directed towards individuals.

We are natural empathetic xenophobes: empathetic by virtue of our intelligence and capacity to apprehend the mental life of a relatively wide range of creatures, but far more narrowly and selectively sympathetic

to the closer group into which are born or choose to join, whereas we tend to remain indifferent or antipathetic to everyone else; neutral or hostile to most aliens.

Judging by present statistics on world poverty, distribution of health care, and the predominantly tense or bellicose relations between individuals, nations, cultures, ethnic groups, social classes, races, genders, religions, political ideologies, etc., the vast majority of human beings appear reluctant or unable to identify with, sympathise or show compassion towards those who are beyond (and sometimes even towards those who are within) “their” sphere. Whilst some societies or individuals may be more prone than others to develop strong ethnic identity, violence, racism, sexism, social hierarchies or exclusion, all exhibit some form and measure of xenophobia.

Thus, in spite of our natural capacity for selective sympathy and mutual assistance that Kropotkin emphasised, the human being also comes very close to Hobbes’ description: a self-interested, control-oriented, fearful, violent, dissociative, conceited, megalomaniac, empathetic xenophobe. In view of their historic prevalence, it is not unlikely that these features have evolved to become a part of our innate neurobiological identity and that any attempt to construe social structures (rules, conventions, contracts, etc.) opposing this identity must, in order to have any degree of realism in application, take this formidable biological challenge into account in addition to the historically well known political, social and cultural challenges. The question can be raised, for example: can – understood as a biological “can” – we develop “global” attitudes (such as the famous first § in the UN Declaration on Human Rights, asserting the equal worth and dignity of all individuals), or are universal declarations doomed to remain mere abstractions because we are neurobiologically conditioned to remain emotionally, and therefore morally, selective and group-oriented? Can sympathy biologically be extended?

The natural egocentricity or individualism of the brain appears quite pronounced: the brain is in constant autonomous activity, projecting autonomously produced images onto its environment that it proceeds to test, and in this activity it refers all experiences to itself, to its own individual perspective. This perspective is naturally narrow, with physical as well as epistemic limitations. We can conceive the narrowness of the individual perspective in terms of space (and the finite perspective’s epistemic limitations) and personal identity (with a typical preference for the self, the familiar, and that with which the individual can identify, to which he or she can relate). Another important aspect of the individual perspective’s narrowness is temporal: it is extremely difficult, sometimes even impossible, for a human being to be emotionally concerned with, or clearly to envisage, actual or possible states or events that are temporally distant (for example, imagined to lie one or several generations ahead in the future) compared to how we are involved with the present. In other words, our cerebral egocentricity is psychological, somatic and spatio-temporal, which means that we, each of us, live in a minute and egocentric world: this-here-now (understanding the “now” as denoting a fairly wide personal time-perspective, since it is notoriously difficult for human beings to live in the “now” understood as the actual present). By nature, we are predisposed to do so: without this massive dissociation we could presumably not survive, at least not with our present cerebral architecture.

A major practical problem is that the effects of our actions are not equally limited. The difficulty of wide-range involvement (be that spatial, temporal or personal) is matched by a facility to cause large-scale destruction on a global scale. This factual tendency to mental myopia that seems to characterise us both culturally and biologically poses serious problems whenever long-term solutions are needed; say, to improve the global environment or reduce global poverty. Our societies are importantly construed around egocentric and short-term perspectives: politically, economically, environmentally, etc., making it extremely difficult to put global or long-term thought and foresight into practice, and this is of course only to be expected if that is the way our brains function.

In this light, it is, we suggest, an important task of neuroscience to diagnose the human predicament in neurobiological terms. What types of social creatures are we, from a neurobiological point of view? Such knowledge can, in addition to its theoretical relevance, be socially very useful and of methodological relevance, e.g., in the development of adequate educational structures and methods, or in the assessment

of alternative methods to remedy social problems. In order to remedy an ill, we first need a proper diagnosis of this “ill”; its nature, underlying causes and theoretically possible remedies. In the absence of such diagnosis we risk opting for methods that may provide a superficial, cosmetic improvement at best, improve appearances perhaps, but without affecting the real situation in any enduring or profound manner.

Importantly, such diagnoses must include both biological and socio-cultural dimensions, as well as a clear understanding of how these perspectives are related. Culture and nature stand in a relationship of symbiosis and mutual causal influence: the architecture of our brains determines who we are and what types of societies we develop, but our social structures also have a strong impact on the brain’s architecture; notably, through the cultural imprints epigenetically stored in our brains. The door to being epigenetically proactive is, accordingly, opened.

3:AM: And that leads me naturally onto my next two questions. Epigenesis is a key area for your thinking. It’s about the way steering the way we evolve by influencing cultural imprints in our brains. Have I got that right? If so, why isn’t this about ‘human enhancement’ and cognitive prosthetics?

KE: The fundamental idea of epigenetic proaction is trying to understand and influence the genesis of human norms in the light of what we today know about the brain. Being epigenetically proactive also means adapting and creating social structures, in both the short and the long term, to constructively interact with the ever-developing neuronal architecture of our brains. It can be described as an educated form of ethical innovation. The scientific challenges involved are accompanied by important social and ethical challenges, some of which we describe below. It seems clear, at least so far as mammals are concerned, that Darwinian evolution has led to the global expansion on the earth of the human species that spends a considerable part of its life developing its brain experiencing and appropriating its physical, social & cultural environment. This is noteworthy: the environmental influence on the brain’s functional architecture yields an evolutionary strength to Homo sapiens, compared to other animals whose cerebral developmental period is comparatively much shorter, and less environmentally determined. That gives an increased importance to “epigenetic mechanisms” driven by interaction with the environment in the course of the long postnatal period of human brain maturation, during which reciprocal relationships grow between the brain and its physical, social, and cultural environments.

Thus our cerebral identity incorporates social interactions. Our brain progressively builds up its connectivity through a constant dialogue between the genetic endowment of the child brain and her/his experiences of the external world. Moreover, trans-generational transfer of information from adult to child takes place through the incorporation of the social and cultural environments in the developing infant brain. From a sociological perspective, these neuroscientific theories and data are important through the scientific support they lend to the idea of adapting social conditions for the benefit of brain functions and their balanced development.

Through epigenetic proaction, new ethical rules can be internally produced and stored in our cerebral architecture. The neuronal features that develop from socio-cultural impact (the results of learning and experience) can be stabilized and passed on through generations, i.e., be epigenetically transmitted. Accordingly, culture can help us in the construction of our brain and conversely, through creative and rational thinking, our brains may as well lead to the production of novel social structures that persist across generations and might be stored in extra-cerebral memories as inscriptions, codes or laws.

In view of this neuro-cultural symbiosis, we can describe this process both as a “neuralization” of the normative process itself, and as a “culturalization” of the brain through the selective stabilization of cultural circuits. Our cultural and social structures – including our normative reasoning – are importantly products of the neuronal structures of our brains, but these neuronal structures are also importantly products of our cultures and societies. Hence the possibility arises to influence our brains with the use of culture, and be epigenetically proactive, in other words: to invent, learn and transmit new ethical norms, forming some kind of new ethical languages.

One of the motivating forces behind our suggestion of epigenetic proaction is a concern about the present state of the world, and the difficulties in dealing with the situation adequately. A similar concern also comes to expression in the current debates on “moral enhancement”.



In these debates, it is often argued that human nature is inapt to handle the problems that humankind presently faces in part as a result of our own actions (environmental destruction, poverty, etc.), and that moral education has not been able to forestall the present global situation described as serious. Another similarity between these two discourses lies in the belief that human nature might be improved, much to the benefit of our societies. The main differences between these approaches lie in the solutions that are suggested.

In the “moral enhancement” debate, the focus is largely on the individual, and the methods suggested are often a kind of “quick-fix” of the brain (such as drugs or brain stimulation) ignoring their short and long term consequences on human brain functions that are under-evaluated and potentially dangerous. In contrast, epigenetic proaction (as I understand it) importantly focuses on the genesis and transmission of novel educational/management programs with long-term influences across generations, and makes no reference at all to direct and blind interventions on the brain. Epigenetic proaction can have important effects on the individual person, and on the individual generation, but it is not conceived as an individual short-cut contrasted to moral education.

3:AM: You discuss this by examining the epigenesis of selective stabilization of synapses. Could you explain what this is and what conclusions you draw?

KE: This part of my work has been inspired by the works of Jean-Pierre Changeux, and my ideas about epigenetic proaction owe much to our long-standing and very fruitful collaboration. During embryonic and postnatal development, the million billion (10¹⁵) synapses that form the human brain network do not assemble like the parts of a computer according to a plan that defines precisely the disposition of all individual components. If this were the case, the slightest error in the instructions for carrying out this program could have catastrophic consequences. On the contrary, the mechanism appears to rely on the progressive setting of robust interneuronal connections through trial-and-error mechanisms that formally resemble an evolutionary process by variation selection. At sensitive periods of brain development, the phenotypic variability of nerve cell distribution and position, as well as the exuberant spreading and the multiple figures of the transiently formed connections originating from the erratic wandering of growth cone behaviour, introduce a maximal diversity of synaptic connections. This variability is then reduced by the selective stabilization of some of the labile contacts and the elimination (or retraction) of the others. The crucial hypothesis of the model is that the evolution of the connective state of each synaptic contact is governed globally, and within a given time window, by the overall message of signals experienced by the cell on which it terminates.

One consequence of this is that particular electrical and chemical spatiotemporal patterns of activity in

developing neuronal networks are liable to be inscribed under the form of defined and stable topologies of connections within the frame of the genetic envelope. In humans, about half of all adult connections are formed after birth at a very fast rate (approximately 2 million synapses every minute in the baby's brain). The nesting of these multiple traces would directly contribute to forming and shaping the micro- and macroscopic architecture of the wiring network of the adult human brain, thus bringing an additional explanation to the above mentioned non-linearity paradox.

Another consequence of the synapse selection model is that the selection of networks with different connective topologies can lead to the same input-output behavioural relationship. This accounts for an important feature of the human brain: the constancy or invariance of defined states of behaviour despite the epigenetic variability between individual brains connectivity.

Finally, both the spontaneous and the evoked activity may contribute to the synapse selection. In this framework, the suggestion was made that reward signals received from the environment may control the developmental evolution of connectivity. In other words, reinforcement learning would modulate the epigenesis of the network. This process of synaptic selection by reward signals may concern the evolution of brain connectivity in single individuals but also exchange of information and shared emotions or rewards between individuals in the social group. It may thus play a critical role in social and cultural evolution.

3:AM: As research moves forward aren't we moving towards the notion of the singularity, the version where we are able to understand the brain to the extent that we can alter it, make it more powerful so we can understand issues that once seemed impossible? And once that happens then what we are doing now and what we are now will seem ultra stupid, in the same way we look at apes. Philosophers and scientists talk about naturalistic responsibility being born out of science's strong social relevance but that's never done much to stop the develop heinous stuff. Aren't there good reasons to fear the advancement of neuroscience – how many sci-fi scenarios do we need to think there could be a problem brewing when superbrains appear?

KE: I think we seem ultra stupid already now, in view of the mess we are making of the world we inhabit. I also think we have good reasons to fear ourselves: our capacities, strengths, and weaknesses. Arthur Koestler compared evolution to a labyrinth of blind alleys and suggested that there is nothing very strange or improbable in the assumption that the human native equipment, though superior to that of any other living species, nevertheless contains some built-in error or deficiency, which predisposes us to self-destruction.

Neuroscience can give us further tools for this destruction as well as help in finding, say, new cures for mental disorders, or better educational methods. But the notion of rapid brain improvement in the form of direct manipulation seems so unrealistic to me that I cannot see any reason to fear it. In spite of the impressive increase of data in neuroscience, our understanding of the brain remains quite modest, and the knowledge we have is not integrated in a manner that would allow us to manipulate it in the way suggested in some SF-movies. That said, I enjoy SF-literature very much, and some of it is scientifically and technologically quite sophisticated.

As for whether or not we should as you query “fear the advancement of neuroscience”, I would reformulate that: it is not the science itself that we should fear but some of the uses that human beings might make of it.

3:AM: Your approach to the brain seems like that of someone like Andy Clark – essentially its ‘an evaluative engine with reward systems engaged in learning and memory as well as higher evaluative tendencies.’ So what do you make of Dave Chalmers challenge that philosophically to explain that we might need to consider matter being conscious all the way down?

KE: Not sure what you have in mind, but I don't see any essential contradiction here? I am favourable to the view that consciousness should be analysed in terms of many levels of cerebral activity – not just the level of conscious awareness, which is quite narrow.

3:AM: And for the readers here at 3:AM, are there five books you could recommend that will take us further into your philosophical world?

KE: There are so many fantastic works to choose between...but selecting a variety from different domains and historical eras, the following five have been rich sources of inspiration to my philosophical thinking in different ways:

1. Epictetus (c. AD 55-135), the [*Enchiridion*](#), or Handbook. Practical wisdom of stoicism.
2. Benedictus Spinoza, [*Ethica*](#) (1677), or Ethics. The most beautiful geometric world-view ever to have been conceived. A forerunner of much modern thought, including monistic views of mind, the concept of libido, and the idea that energy constitutes matter.
3. Immanuel Kant, *Kritik der reinen Vernunft* (1781/87), or [*Critique of Pure Reason*](#). A fundamentally logical Weltanschauung that remains relevant to contemporary philosophy.
4. Arthur Koestler, [*The Ghost in the Machine*](#) (Hutchinson, 1967). A thought-provoking philosophical-psychological treatise on human nature and human tendencies towards self-destruction.
5. Jean-Pierre Changeux, *L'Homme de vérité* (Odile Jacob, 2002), or [*Physiology of Truth*](#) (Harvard University Press, 2004). A neuroscientific work of great philosophical importance about how the brain seeks knowledge.



ABOUT THE INTERVIEWER

[Richard Marshall](#) is still biding his time.

Buy [his book here](#) to keep him biding!



Most Recent

- [Interviews » The End Times » Neuroethics](#)
- [Poetry » The Enemies project 2015: a year in review](#)
- [Interviews » The Indie Press Interviews 2: Jacques Testard](#)
- [Reviews » After something else: A review of Emblems of the Passing World by Adam Kirsch](#)
- [Essays » Masturbating in Public; or Misreading Kafka](#)
- [Fiction » Tolly](#)
- [Reviews » I can only speak for myself](#)
- [Reviews » A Book Review That is Not One: On Maggie Nelson's *The Argonauts*](#)
- [Reviews » Eyes](#)
- [Essays » The Exploitation of the Technical](#)

• **Buzzwords**

- [Top Reads of 2015: Andrew Stevens](#)
- [Top Reads of 2015: Tristan Foster](#)
- [Top Reads of 2015: Joanna Walsh](#)
- [Top Reads of 2015: Dustin Illingworth](#)
- [Top Reads of 2015: Fernando Sdrigotti](#)

- [Top Reads of 2015: Sam Stolton](#)
 - [Top Reads of 2015: K. Thomas Kahn](#)
 - [The Missing Links](#)
 - [Tonight in London: A Beast in View](#)
 - [my brooklyn writer friend](#)
-

All content © 2000-2015 / Designed and built by [Rhys Tranter](#) and [Florian Kräutli](#). Based on an earlier structure built by [STML](#).

Email webmaster@3ammagazine.com with any problems.

[Login](#)