

Julian Nida-Rümelin

**Is there a neurophysiological challenge
to freedom and responsibility of human action ?**

HARRIS-LECTURE ON SCIENCE AND CIVILIZATION

March 3, 2005

California Institute of Technology

Pasadena, CA

Dear Mrs. and Mr. Harris

I am glad that the Harris-Family who sponsor this *lecture on science and civilization* has come from San Francisco to listen my talk

dear host Prof. Diana Buchwald whose caring mind is concerned with every detail of our stay and made me, my wife Nathalie and above all our 17 months old daughter feel at home from the first evening after our arrival

dear colleagues from CALTECH one of the world's most reputed academic institutions, which practices such an open-minded and warm-hearted hospitality to visiting researchers faculty and students

ladies and gentlemen,

I am going to tackle a difficult topic in this lecture a topic which is almost as old as philosophy itself. I am going to do this in four steps

First I will give an account of how we should understand *freedom of human action*. This account – you may call it a Kantian or Stoicist account - is based on a close relation between freedom and reason. Therefore the first part of my lecture has the title “Freedom and Reason”.

Second I will discuss how responsibility comes into play here and we will see that there is a complex of moral attitudes which is constitutive for the human condition in which responsibility and freedom are closely interconnected. Therefore the second part of my lecture has the title “Freedom and Responsibility”.

Third I will answer the question whether there is a neurophysiological challenge to this humanist account, that we developed in part one and two. Therefore the third part of my lecture has the same title as my talk: Is there a neurophysiological challenge to freedom and responsibility of human action ? The answer will be “no” notwithstanding the impressive empirical findings of recent neuroscientific research.

Since I should present my arguments in less than an hour, and since I think quickly but speak slowly even in German, much more though in English, many questions will remain open and therefore I am glad to hear that there will be time for discussion after my talk. May I add that within some days a booklet on this topic comes out under the title “Über menschliche Freiheit”, published by Reclam – for those who can read German.

I

Freedom and Reason

Since Greek antiquity it was a challenge to philosophical and everyday thinking alike to reconcile the assumed determinism of the natural world with freedom of human action. The Stoicists discussed this question at length and some historians of philosophy think that the Stoa as a forceful cultural movement collapsed because the Stoicists did not succeed in developing a convincing solution for this problem. Immanuel Kant developed his *Critique of Practical Reason* in order to answer the question of how human freedom can ever exist in a world governed by determinist laws. His answer was roughly that this can only be understood, if man has a specific capacity to give himself the laws that govern his actions, in short: *autonomy* is his answer.

The problem with Kant’s ‘solution’ is its dualism. Man as part of nature is governed by causal deterministic laws like everything else in nature. Man as *noumenal* entity is not governed by the laws of nature but by the laws he gives himself based on his rational capacity. But are both simultaneously conceivable? *Compatibilists* say “yes, determinism and freedom are compatible”, the *Incompatibilists* say, “no, they are not”. I think that some form of

incompatibilism is basically right and in this respect I agree with the common man, but with only few philosophers. The common philosopher adheres to compatibilism. Libertarians assume that man is free and not be subjected to deterministic causal laws. Libertarians are incompatibilists, but most incompatibilists are anti-libertarian. They think that the natural world is causally closed: that the natural world is without any residuum governed by the laws of nature, that man is part of this natural world and that freedom and responsibility in the sense understood by the common man and some philosophers are illusions. Some neuroscientists say that since recently there is empirical evidence that the human brain shows all characteristic traits of a deterministic system governed by neurophysiological laws, that what seems to be a motivation or an intention or a reason to act is in fact *ex post*, is an interpretation, which the human brain delivers after the action was already causally initiated. There are experiments, the most famous are those of Benjamin Libet, which seem to affirm that.

I think Kant – and the Stoicists – were on the right track. It is reason - the capacity to reasoning and the effectiveness of reasons for judging and acting – which constitutes human freedom. Man's action is free insofar as it is guided by reasons and is not merely the causal effect of natural facts and events. In other words: *liberty is the naturalist underdetermination of reasons*. Reasons play a causal role for our behavior, but reasons themselves are not part of nature's lawful empire. If every element of our behavior could be causally explained using the conceptual means of physics, chemistry, biology, neurophysiology or other natural sciences alone, *practical reasons* would not be relevant for, what we do. And likewise if, all our beliefs could be causally explained using the conceptual means of physics, chemistry, biology, neurophysiology or other natural sciences alone, *theoretical reasons* would not be relevant for, what we believe. We might then still have the impression that reasoning played an important role for what we believe and do, but this impression would not be justified, it would be an illusion. Some neuroscientists think that in this sense we are caught by an illusionary self-image.

This account of reasons differs radically from others, e.g. that, which takes arbitrariness as the core of human freedom. Think of a person, who has a strong character guided by moral convictions. She decides not to blame somebody for some wrong decision taken, although to blame him would make the job much easier for her. Asked why she didn't she answers, "well I couldn't because I wanted to be fair with a man new on this job". She said "I could not" in the sense that if she did, she would have acted against some firm moral attitude she has. It was far from arbitrary that she decided as she did. There was no other alternative open for her, as

the person she is. But even if we say that there was no other alternative for her, as to act such as she did, we would not say that her action was fixed in advance independent from her reasoning. She thought it to be unfair to blame him and she took that as a good reason not to blame him. This reason guided her to act as she did. If there were causal constraints on her action, if for example there was no physical possibility to blame him, and if this fact caused her not to blame him, then her action would not have been guided by reason, it would not be a free action. It is reason which constitutes freedom.

II

Freedom and Responsibility

For a long time, a largely accepted view in philosophy regarding human freedom was given by Moritz Schlick: [...I must speak [in this chapter] of a matter which is still considered to be a fundamental question of ethics, but which is only made into one through a misunderstanding in ethics: the so-called problem of freedom of will and he adds, “it is really one of the greatest scandals in philosophy that so much paper and ink is wasted on this stuff.”¹ One might add: “...and not only at that time, but especially now, once again, at the dawn of the 21st century, and precisely in the disciplines of neuroscience and analytic philosophy.”

Getting rid of this philosophical scandal is quite simple, according to Schlick. One must merely follow two views: first, that ethics is not about prescribing, but rather about explaining human behaviour, and secondly, that ethics assumes the causal principle in virtue of being a nomologic science. Furthermore, ethics has to do with *responsibility*, and this can be shown in the fact that behaviour can be influenced by *sanctions*; hence, causal determination is not only no threat to human responsibility, it even makes responsibility possible at all.

In a famous article “Freedom and Resentment”² (1962, more than 30 years later) Peter Strawson’s reply consisted of two parts:

(1.) Our moral emotions and attitudes are expressions of complex relationships between persons. Unlike Schlick, Strawson was less interested in moral practices such as punishment, moral condemnation, or moral approval than what he referred to as ‘non-detached attitudes and reactions of offended parties and beneficiaries.’ He discussed in

¹ Cf. Moritz Schlick: *Fragen der Ethik*, Wien 1930, chap. 7

² In: *Proceedings of the British Academy* 48 (1962).

detail resentment, gratitude, and forgiveness. Strawson's argues convincingly that moral emotions and attitudes were not effective, if they were only used instrumentally as a means to influence, control and manipulate behaviour.

(2.) It is practically impossible for a theoretical conviction—such as the belief that every event in the world is causally determined—to violate our moral intuitions so deeply that the interpersonal relationships, in which they find expression and are trusted, fail to survive. Mentally sane human beings can not adopt an attitude towards interpersonal relationships' deprived of genuine moral sentiments, because the isolation that would result from it transcends human capabilities.

Our interpersonal relationships, which find expression in our everyday interactions, and more specifically, in our moral attitudes, assume that humans are *responsible* for their actions: that they are not merely objects of causal influence (whether it be physical, biological, neurophysiological, or psychological). As normal human beings, we are embedded in social relationships and ascribe responsibility and freedom as is required of the moral emotions and attitudes shared by us all. The experience of our everyday interpersonal relationships does not leave any room for a theoretical conviction that would make these attitudes disappear.

We develop moral attitudes and emotions under the assumption that we are the *initiators of our actions*—i.e. that we act freely and responsibly. The perspective changes when we are confronted with borderline cases: for example little infants or the mentally ill are not, or are only to a lesser degree, accountable for their actions. It would be inadequate to take an utterance which is normally insulting, as such. In cases such as these, we no longer appeal to good reasons for or against, we simply exhibit the behaviour that we hope will bring about the desired effect (for example, deterrence or provocation). The attitude on which such behaviour is based would be inappropriate if directed toward ordinary persons—it would if adopted generally destroy the essential traits of our life-world interactions.

The reactive moral attitudes, based on the ascription of responsibilities would appear inadequate to us if human behaviour, actions and judgements were not free—regardless of whether they were determined by causal laws or contingently through accidental processes. But 'inadequate' is certainly not the same thing as 'psychologically impossible'. A moral attitude can be inadequate under the assumption that the given person has a certain (descriptive) conviction, and still there may be individuals who have both this conviction and this attitude. In view of this combination of moral attitude and descriptive conviction, we consider such individuals to be irrational.

The decisive question is whether or not a person's behaviour is controlled by reasons; whether the person is able to weigh reasons for and against; whether objections and counter-reasons are accessible to the person—in short, whether one can assume a sufficient degree of rationality. By 'rationality', I mean *acting and judging on the basis of reasons*. The rational person can provide coherent explanation in terms of practical and theoretical reasons, when asked why she did or believed this instead of something else.

Insofar as we allow (or grant) that a person's actions are guided by reasons, i.e. we hold her responsible, approval and disapproval as well as feelings of resentment or offence may be appropriate. However, these and other moral attitudes and emotions require that reasons not be merely subjective. I can only blame someone for something if it is possible to have different opinions about whether or not something is an appropriate reason, and if reasons aren't arbitrary: if there are criteria which distinguish good reasons from bad ones. The morality of our »Lebenswelt« requires *fair* approval and disapproval, which means that one can't simply take his own personal interests as a standard for judgement; it requires a certain degree of *objectivity*. We dispute the appropriateness of reasons; we are certain that some reasons are appropriate while others are not, but this only means that we are not at liberty to just choose the reason best suited. Strawson points to the tension between the active (partaking) mindset and the objective mindset, and he mentions that we could even consider this to be the tension between our humanity and our intelligence. I think a different accentuation and a different conceptualization are necessary, if we are to properly understand our human condition.

In order to clarify this new accentuation further, let us consider an example that is already quite familiar to many of us. An elderly person is losing more and more of his memory capacity. At the onset of this process, there are always disagreements brought about due to this loss. The elderly man, becoming more and more forgetful, is firmly convinced that his memory is not deceiving him. He lives with this conviction in the expectation that nobody will contradict him. The younger family members answer back with arguments that should demonstrate how this memory *must* be false, and so forth. A few years later—the amnesia having advanced in the meantime—there are no more arguments. Such arguments are kept to oneself as one carefully manoeuvres so as to avoid such disputes and to maintain as much familiarity as possible in the home. This does not even serve as a description of moving from a subjective to an objective mindset, let alone of a transition from humanity to intelligence. A different, more rigid, more demanding, countenance is needed for interacting with individuals who are perfectly sane and mentally fit. Criminal Law is quite clear on this point. Any

criminal who was certifiably sane at the time of committing the crime, must be held fully accountable or responsible for what he has done. Whoever's actions are guided by the weighing of reasons must be able to stand by his actions. Humanity—in the sense of considerateness and forgiveness, cautiousness and empathy—must be shown especially toward those who are not fully in control of their lives.

And *vice versa*: whenever we weigh our reasons, when we check whether a certain behaviour is morally acceptable or not, we always make use of shared standards, of rational arguments. We don't have to imagine ourselves in the position of others to know that a murder or an insult is morally unacceptable. The sad biographical stories told in court as an excuse for the accused are only relevant if the accused, due to the related psychological disorder, was unable to weigh his reasons and act accordingly. Whether or not a punishment is appropriate is decided on the basis of specific characteristics of the action, as they are described over the course of the trial. The judicial community has decided to declare certain characteristics to be crucial or decisive. They have an objective, or at least inter-subjective, character, and we are convinced that they are normatively appropriate, by and large. Legal punishment is a form of institutionally governed social resentment. Early parole for good behaviour is a form of social forgiveness. Hence, just as punishment and early parole fulfil, or should fulfil, certain criteria, so too are everyday cases of resentment and forgiveness, to a large extent, rooted in shared and discussible criteria of adequacy. If I resent someone, I should have reasons for doing so. So I must be able to show that the person acted wrongly, in order to be justified in my resentment toward him. Forgiveness also has its limits. Not everything can be forgiven on a whim of good humour: we generally expect good reasons here, too, but the difference must be admitted between everyday acts of forgiveness, as (informal) pardons, and the legally guided processes involved in granting early parole for good behaviour.

We couldn't view ourselves and others any other way than as beings who orient their actions and judgements around reasons; as beings who have reasons for their actions and judgements which they could articulate, if queried, who are free and responsible insofar as they judge and act guided by reasons.

This perspective does not apply to every human being, rather only to those who possess certain (minimum) cognitive and moral abilities, in short: who are *responsible* subjects. The ability to assume this, both in ourselves and in the people we talk to and interact with, underpins the entire spectrum of reactive moral attitudes and sentiments. We cannot remove this assumption without collapsing the whole system. The fundamental (and, if you like,

‘transcendental’) rôle of this premise of rationality, freedom and responsibility renders naturalistic reinterpretation strangely presumptuous as the expression of an attitude which disavows that we’re not talking about a corollary of a scientific worldview, but rather about the composure of the interactions in our »Lebenswelt«, *in toto*.

III

Is there a neurophysiological challenge to freedom and responsibility of human action ?

The progress of neuroscience is impressive and it promises to deliver a much deeper understanding of the material basis of human emotion, reason, and decision-making. Those neuroscientists who are involved in the most ‘cutting-edge’ research, however, stress that, despite impressive successes in the field, many brain processes still remain unexplained. Like other successful scientific disciplines of earlier eras, neuro-physiology is accompanied by a worldview, which goes far beyond the discipline and its specific competence. Just as the universal Determinism of the 18th century accompanied classical Physics or as the (socio-political) ‘Social Darwinism’ of the 19th century arose from the Darwinian revolution in Biology—so too, I expect, will the present phase of a worldview inspired by Neuro-physiology soon wane in response to increasingly complex experimental results. It is nevertheless sensible to give warning at an early stage of such a “neuro-ideology” and I suppose that such a warning fits well in a *lecture on science and civilization*.

The current public debate about free will between neuroscience and philosophy gives the impression that two faculties stood in opposition to one another: the natural sciences, especially the neuro-science, on the one hand, and the humanities, including philosophy on the other. The one faculty allegedly has evidence that freedom of will is an illusion, the other refuses to accept these insights, preferring to maintain an age-old European conception of man which one may find endearing, but which has already been surpassed. The one faculty employs empirical arguments, while the other is allegedly bound to a Cartesian view of mind that can no longer be upheld in light of the knowledge that we have gained from neuro-physiology.

This opposition represents an inaccurate portrayal of the actual scientific state of affairs. E.g. I developed my own position of ‘theoretical humanism’ in opposition to the dominating

compatibilist philosophical arguments of past decades. Most philosophers who have addressed this issue in the 20th century have tried to make human freedom into something that seamlessly fits a conception of the world governed by causal deterministic (or probabilistic) laws. The Cartesianism that is attributed by many neuroscientists to philosophers in general only accounts for a small minority of contemporary philosophical positions.

The Libertarians are non-compatibilists; this means they hold the assumption that the course of events in the world is completely determined (by deterministic laws) to be irreconcilable with the existence of human freedom and they additionally believe that there *is* human freedom. Very few Libertarians are Cartesians, i.e. very few assume that there are material things (*res extensae*) on the one hand and mental things (*res cogitantes*) on the other, which exist independently from each other. Cartesianism and Libertarianism are often confounded in the current debate. Libertarianism claims that complete determination of the course of events in the world is incompatible with human freedom, and that there is such a thing as human freedom. It is implicit in this claim, however, that complete determination of the course of events is not completely laid down by deterministic laws of nature. The deterministic, or better yet ‘naturalistic’, school of non-compatibilism claims the opposite: on the one hand, it is in agreement with the Libertarians that complete determination of the world is irreconcilable with freedom of will, while on the other hand, it denies human freedom and responsibility. On this view, the course of events is taken to be completely set down by deterministic and probabilistic laws of nature, and accordingly, human freedom is excluded. Those neurophysiologists who reject human freedom accept a naturalist and non-compatibilist position.

The current debate about the existence of human freedom and responsibility is plagued by the fact that it generally remains unclear how such a thesis could be empirically proven or disproven at all. Neurophysiologists claim, for example, that there is no ‘control centre’ in the human brain. Philosophers have rightly answered that no such *homunculus* conception has been held in philosophy and that, in addition, this empirical discovery is inconsequential to the question of freedom of will. If our argumentation is correct, then such a ‘proof’ [that there is no freedom of will] would have to demonstrate the completeness of the neurophysiological description: it had to be shown through this empirical discovery that every action and decision, every deliberate and emotional process, could without any residuum be described, explained, and predicted without reference to reasons by exclusive use of neuro-physiological concepts and laws. It is certainly a good heuristic principle for neuroscientists to assume that they could, but this in no way proves this assumption to be correct.

As a proof of the claim that there is no freedom of will, many neurophysiologists have proposed that the human brain is a strictly determined system such that the probabilism of the microphysical level is no more pervasive than chaotic processes.³ It seems to me that it is still much too early for such a claim. No one can say, at this point, whether micro-physical indeterminism plays a macro-physical role in the brain, and whether the strengthening mechanisms known from chaos theory play a role in mental activity. The complete and naturalistic determination of temporal courses of events of the brain state could not be satisfied until it were shown that one's genetic make-up, together with the effects of one's environment, would be enough to uniquely determine the brain state of the subject at any given time. Any respectable scientist would certainly grant that such a proof, given our current scientific understanding, is nowhere near completion.

One peculiar and particularly misguided neurophysiological train of arguments consists in using an intentionalistic language in order to describe physiological brain states. It is said e.g. that the brain 'strives toward consistent states', 'tries to resolve tensions', 'to coordinate operative processes', and to provide 'coherent interpretations' for the vast multitude of sensory data. The brain is taken to use 'the temporal dimension as a coordinating space', and uses 'precise temporal synchronization as a code for grouping neural answers', etc.⁴ It should be clear that the question of whether or not weighing reasons plays a causal role for our beliefs and actions cannot be answered by maskering the physiological description as complex intentional processes of the subject's brain. If mental events and processes are realized in *neuro-physiological* events and processes, the discovery of this correspondence does not, of course, disprove the causal role of intentionality or deliberation. The various arguments brought forth are only suited to affect those positions which, in the Cartesian tradition, claim that mental and especially deliberative processes are completely independent from physiological brain processes. However, I, for my part, do not know of a single contemporary philosopher who holds such a view. At least I can assure you that I do not hold such a few.

In support of the claim that there are no such things as human freedom and responsibility, some neurophysiologists say that animal behaviour is completely determined, that the corresponding actions which follow are a necessary result of the interrelation between the

³ Cf. P. Smith-Churchland: *Brain-Wise. Studies in Neurophilosophy*, Cambridge 2002.

⁴ Cf. Wolf Singer's „Selbsterfahrung und neurobiologische Fremdzuschreibung. Zwei konfliktrichtige Erkenntnisquellen“, MPI Hirnforschung, Frankfurt am Main 2004 and id. „Consciousness and the Binding Problem“, in: *New York Academy of Sciences* 929 (2001), p. 123-146.

subject's current state of sensory stimulation and the brain state prior to this stimulation. On the other hand, however, the neuroscientists give proofs that the brains of humans and those of animals are almost identical, and that their development, their construction, and their functions obey the same principles. The claim is that since we have no reason to doubt, with regard to the brains of animals, that behaviour is always based on neural functions and that it is therefore subject to deterministic laws of physical and chemical processes, this claim of material conditionality must also apply to humans. This argument is misleading insofar as it is always a requirement when ascribing responsibility to a person, that this be done only under circumstances that appear appropriate. Infants can never be held responsible for their actions; drunks, at most, for their having gotten drunk; and victims of psychological trauma are held responsible to a lesser degree because we (in these and other such cases) doubt whether the person involved was able to weigh reasons and act accordingly. It is certainly only a matter of degrees between being more or less drunk; however, this range of comparison is exceeded by the degrees separating humans from rhesus monkeys. Thus we are gradualists in our everyday ascriptions of rationality, freedom, and responsibility: rationality, freedom, and responsibility are matters of degree and then the question arises whether we really 'know' that the brains of highly developed mammals are strictly determined systems. Behavioural biologists have collected extensive data that seems, rather, to speak against this assumption.

Most neurophysiologists identify themselves with some form of *epistemic constructivism*, according to which we have no access to reality but are rather dependent, in our perceptions and beliefs, on the optimally suited sensory apparatuses that have resulted from the process of evolution. Of course, one might immediately ask whether the justification of constructivism is self-affirming since some empirical data is introduced in this justification, the dependability and objectivity of which is simply assumed in advance. But apart from the self-contradictions inherent in constructivism, most neurophysiologists tend to accept an incoherent position, which identifies itself with constructivism on the meta-theoretical level, while at the same time remaining bound to a naïve form of realism in their scientific work. Anyone who takes part in the justificatory game of science (to put it in Wittgensteinian terms) at least implicitly accepts that reasons play a role, that there are good and bad reasons for assuming certain things: for example, that there is human freedom. If these reasons were nothing more than *ex post* rationalizations of causally determined brain processes, then the exchange of arguments for and against would be pointless; it would be irrelevant.

Reductionistic metaphysics plays an important part in the claim that there is no freedom of will: it is assumed that all psychological processes are nothing else than neurophysiological

processes, and that the latter are self-contained. This is not the appropriate place to go into the difficult problems concerning reduction in scientific theories⁵, a caveat is nevertheless in order. The reduction of all of natural science's laws and descriptions to physics, as scientists around the beginning of the last century still thought to be close at hand, hasn't made much progress: on the contrary, the multiplicity of nomologic descriptions has grown, just like the zoo of elementary particles. In previous years considerable progress has been made in the field of probabilistic theory of causality, which dissolves the concept of causality from deterministic laws. This development suggests that it is advisable to be careful in handling the concept of causality and to give up on linking determinism and causality. The assumption that any given brain state is completely determined by genetically programmed organization, by epigenetic influences, and by environmental influences in combination with what happened directly prior to the moment under consideration, seems to me—in view of all of this—to be based more on speculative metaphysics than empirical discovery. The prognostic ability of neurophysiology is, to date, still much less definitive than that of our everyday psychology. Predicting the behaviour of humans on the basis of their brain states and environmental conditions will remain (in the foreseeable future) more difficult than making behavioural predictions on the basis of experience gathered in everyday life and through our orienting knowledge, which belongs to what we refer to in English as *folk psychology*. The natural sciences, neuroscience included, are at present still a long way from being able to explain and predict our deliberations, emotional composure, and acting over time. Since human freedom is nothing else than naturalistic under-determination of reasons for action and judgement (assumed that our argument is valid), it cannot yet be regarded as having been refuted, given the current stage of neuro-science and our moral attitudes and sentiments, our form of everyday interaction, make it highly implausible that there will ever be such a refutation.

Prof. Dr. Julian Nida-Rümelin
Visiting Associate at the Division of Humanities and Social Sciences
California Institute of Technology
Pasadena, USA

Chair for Political Theory and Philosophy
Geschwister-Scholl-Institut
Universität München
Germany

Julian.Nida-Ruemelin@lrz.uni-muenchen.de

⁵ I explain my position in „Reduktionismus und Holismus“, in: *Technomorphe Organismuskonzepte*, ed by W. Maier und Th. Zoglauer, Stuttgart-Cannstadt 1994, pp. 25-46.

