frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 14 March 2012
doi: 10.3389/fpsyg.2012.00067

On some unwarranted tacit assumptions in cognitive neuroscience†

Rainer Mausfeld*

Department of Psychology, Christian-Albrechts-Universität zu Kiel, Institut für Psychologie, Kiel, Germany

Edited by:
Bernhard Hommel, Leiden University, Netherlands

Reviewed by:
Lorenza S. Colzato, Leiden University, Netherlands
Bernhard Hommel, Leiden University, Netherlands

****Correspondence:***
Rainer Mausfeld, Department of Psychology, Christian-Albrechts-Universität zu Kiel, Institut für Psychologie, Kiel D-24098, Germany. e-mail: mausfeld@psychologie.uni-kiel.de

The cognitive neurosciences are based on the idea that the level of neurons or neural networks constitutes a privileged level of analysis for the explanation of mental phenomena. This paper brings to mind several arguments to the effect that this presumption is ill-conceived and unwarranted in light of what is currently understood about the physical principles underlying mental achievements. It then scrutinizes the question why such conceptions are nevertheless currently prevailing in many areas of psychology. The paper argues that corresponding conceptions are rooted in four different aspects of our common-sense conception of mental phenomena and their explanation, which are illegitimately transferred to scientific enquiry. These four aspects pertain to the notion of explanation, to conceptions about which mental phenomena are singled out for enquiry, to an inductivist epistemology, and, in the wake of behavioristic conceptions, to a bias favoring investigations of input–output relations at the expense of enquiries into internal principles. To the extent that the cognitive neurosciences methodologically adhere to these tacit assumptions, they are prone to turn into a largely a-theoretical and data-driven endeavor while at the same time enhancing the prospects for receiving widespread public appreciation of their empirical findings.

Keyword: cognitive neuroscience

The cognitive neurosciences originated about three decades ago, instigated by the invention of new technologies for the study of brain processes. Since their proclamation they have attained a triumphal success, both in psychology and in the general public. The sheer quantity of corresponding research alone bears witness to this success. It is a reasonable guess to assume that presently more scientists, in the broad spectrum of fields that goes under the label of cognitive neurosciences, are studying the foundations of mental phenomena than in the entire 2000 years before. Measured by criteria such as the number of scientific disciplines involved and of researchers participating, or the quantity of published papers, or acquired grant money, the cognitive neurosciences undoubtedly are the most successful enterprise in the entire history of psychology.

The cognitive neurosciences are a highly variegated interdisciplinary field involving a great range of disciplines, stretching from genetics, biophysics, and neurophysiology to computational approaches to psychological phenomena. Despite these differences, these fields share a defining feature by which they regard themselves specifically as cognitive neuroscience. This is the guiding idea that reference to the behavior of neurons or populations of neurons enters essentially into any explanatory account of psychological phenomena and achievements. Accordingly, the cognitive neurosciences focus on investigations of the potential neural substrate of mental processes.

The corresponding experimental investigations have brought forth a wealth of intriguing new findings concerning the relation of mental phenomena to neural processes and have advanced neurophysiological understanding on all levels of analysis. The observations and findings that reveal systematic correlations between certain kinds of mental phenomena and specific neural events in the brain have found great resonance in the general public, and have given rise to the expectation that a breakthrough in a deeper understanding of mental phenomena is at hand. Accordingly, the cognitive neurosciences are highly successful in attracting public attention and became an immensely popular programmatic research perspective. For a field of foundational research, this public resonance is quite remarkable and hardly a matter of course. Rather, the history of science mostly testifies to the contrary (alchemy being an outstanding exception). The question then arises as to the factors that made this unique public success possible. The cognitive neurosciences are arousing a public enthusiasm that could scarcely be the effect of achievements pertaining to a deeper theoretical understanding. Advances in theoretical understanding tend to be tied, in the natural sciences, to reformulations and radical re-conceptualizations of our ordinary modes of thinking, and hence go along with an increasing tension to our common-sense intuitions. Thus, serious theoretical advances will, as a rule of thumb, not elicit much enthusiasm from the broader public. Of course, in the case of theoretically mature disciplines, such a public success could result from developments that yield new technologies with a high impact on society, or that challenge deep-seated philosophical world-views. But how can a field of foundational research that deals with phenomena at the very boundary of theoretical understanding exert such an overwhelming influence in almost all fields of psychology and

†This paper is partly based on Mausfeld (2010b).

receive such widespread and enthusiastic resonance in the media and the general public?

The reasons can hardly lie in theoretical breakthroughs with respect to our understanding of mental phenomena. By the standards of theoretical explanation that have been established in the development of the natural sciences, our theoretical understanding of even the simplest mental phenomena still barely scratches the surface. Take for instance visual perception, arguably the oldest and theoretically most mature field of systematic psychology enquiries. Nowhere in psychology should the prospect be better to identify neural foundations of psychological phenomena. Indeed, most standard text book chapters on visual perception convey the impression that current neurophysiology offers explanations for a variety of basic phenomena, and perceptual psychologists are notoriously inclined to hastily call upon ad hoc pseudo-explanations for isolated psychophysical phenomena in terms of equally isolated neurophysiological findings. As a matter of fact, however, such explanatory claims are unwarranted. The discrepancy between the explanatory value of neural correlates and their actual contribution to a better theoretical understanding of basic internal principles underlying mental phenomena becomes all the more discernible the better our theoretical understanding already is. Color perception, which provides a comparatively rich body of theoretical substratum, is an instructive case in point. With respect to primary color coding, as captured by the Grassmann laws of metameric color matches, the prospects for linking psychological phenomena to neural events should be exceedingly promising. Such an assessment, however, turns out to be utterly inappropriate. For instance, McLeod observes that not even the most elementary properties such as trichomasy are presently understood in their neural foundations and that even for elementary perceptual relations "the significance of neural events becomes increasingly obscure." (MacLeod, 2010, p. 172/173) With respect to a different perceptual achievement, Zeki (1998) noted that the perception of a "continuous line is a mystery that neurology has not yet solved." As to the claim that a theoretical understanding of visual perception derives from neurophysiological investigations, Barlow (1983, p. 11) emphasized: "Nothing could be more misleading, for all the important properties of the visual system were first established by psychophysical and psychological observations made on the system working as a whole. [. . .] physiologists need to be told what the visual system does before they can set about the difficult task of finding out how it does it." Accordingly, advances in our psychological understanding of perceptual phenomena have in the first place benefited and fostered neurophysiology rather the other way round.

When we turn to most fundamental principles underlying perception – such as the nature of "perceptual objects," or, more generally, the nature of data formats or conceptual forms on which any kind of perceptual (and cognitive) information processing by definition has to be based –, it becomes entirely opaque how neurophysiological findings could contribute to a deeper theoretical understanding of these foundational aspects. Among many other deep theoretical challenges, perceptual psychology still lacks an appropriate theoretical understanding of its core notion, viz. that of a "perceptual object" (cf. Mausfeld, 2010a). Corresponding theoretical challenges and problems have neither been alleviated nor even been touched by empirical findings from cognitive neuroscience.

Given the present state of our theoretical understanding of mental phenomena and achievements, the overwhelming influence of the cognitive neurosciences cannot be attributed to theoretical breakthroughs by which we have gained some deeper understanding. All the more as such an understanding will most likely lie on a level of abstraction that is far beyond our ordinary modes of thinking and hence will not be well-suited to yield widespread public resonance. In contrast, empirical findings of neural correlates of psychological phenomena such as depression, cooperative behavior, empathy, political attitudes, or religious feelings conform to core features of ordinary modes of thinking or can be accommodated to them. Hence, they are suited to attract the attention of the general public.

We therefore have to look for factors outside the realm of developments of explanatory theories in order to better understand the dominance that the cognitive neurosciences currently exert over almost all fields of psychology. This dominance is crucially due, or so I will argue, to the impact of certain common-sense conceptions[1] on central but mostly tacit assumptions underlying the cognitive neurosciences. The detrimental impact of common-sense intuitions on the development of scientific theories has been a matter of extensive enquiries in the history and philosophy of science. With respect to the cognitive neurosciences, this impact would merit closer scrutiny and deeper discussion. I will, however, confine myself here to providing a few reminders of what is or ought to be regarded as methodological truisms in the natural sciences.

In order to prevent potential misunderstandings as to my intentions and the arguments adduced, I wish to emphasize a few points from the outset. (i) My remarks will focus on specific issues that arise in the context of efforts toward theoretical understanding and explanation in *psychology* (understood as part of the natural sciences). My concerns therefore relate primarily to those areas of the cognitive neurosciences that deal with foundational aspects of psychology and aim for a deeper theoretical understanding of mental phenomena. Areas of the cognitive neurosciences that are first and foremost driven by neuroscientific purposes, or that focus on animal studies are naturally less vulnerable to the adverse influences of the kind of tacit common-sense intuitions addressed here. Also, applied research in the cognitive neurosciences is not a target of my concerns here, because it follows its own proprietary

[1] I will understand, in the present context, the terms common-sense conceptions, everyday conceptions, or ordinary modes of thinking in the broadest possible way, namely as the diversity of modes by which we conceive of psychological phenomena in all contexts other than that of the natural sciences. This usage comprises not only those concepts and ways of world-making, which underlie, as part of our biological endowment, our ordinary discourse about the world and our acts of perceiving, but also derived concepts and notions that have been developed for other purposes than those of the natural sciences, whether technological, philosophical, or of any other kind. The distinctive feature of common-sense conceptions is their intuitive plausibility, whereas conceptions in the natural sciences receive their value first and foremost from their explanatory power. Our common-sense categorizations of phenomena and our common–sense conceptions of explanation are part of our conceptual endowment (notwithstanding that they are multifariously molded by cultural processes), and hence constitute themselves an important object of enquiry of cognitive science.

lines of thinking and employs criteria of success different from the ones underlying foundational research. (ii) The problems and issues to which I want to draw attention are well-know from the history of the natural sciences and have been amply discussed in great depth in the corresponding literature. They pertain to core methodological aspects in the development of scientific theories, and hence cannot be demoted to being merely philosophical ones. Accordingly, I am solely interested here in some very general methodological issues that are intrinsic to the development of explanatory frameworks in the natural sciences. I have no interest in issues that, in the context of scientific enquiries, could be discarded as belonging to philosophy. (iii) These notes are not meant to be an evaluation or a review of the present status of the cognitive neurosciences; they merely have the goal of drawing attention to some theoretical distortions of currently prevailing approaches. Furthermore, they inevitably cannot do justice to the intricacies and subtleties that are intrinsic to the issues addressed; rather I have to confine myself to a treatment on a level of abstraction on which most of the claims made should be rather uncontroversial in the context of the natural sciences. (iv) It is not my intention to question in principle the integrative purposes of the cognitive neuroscience. To repeat, the cognitive neurosciences have yielded a wealth of intriguing new findings concerning the relation of mental phenomena to neural processes and have advanced neurophysiological understanding on all levels of analysis. My intention rather is to point out that currently prevailing approaches in cognitive neuroscience are at risk to violate core methodological principles of the natural sciences because they are based, in a deleterious way, on tacit common-sense assumptions.

Before I address the tacit common-sense assumption involved, I will start with some brief preparatory remarks on the constitutive characteristics of the cognitive neurosciences, namely their focus on the neural foundations of mental phenomena.

MENTAL PHENOMENA AND THE QUEST FOR THEIR "MATERIAL" FOUNDATIONS

Mental phenomena are part of nature and hence belong to the types of phenomena that we try to integrate, in natural science, into a coherent theoretical framework. Attempts to theoretically understand the nature of mental phenomena proceed on the general assumption that mental phenomena are brought about by specific properties of a biological system, namely the brain. The corresponding premise has been regarded as a natural, basic assumption of science since the 18th century at the latest (while more recently – and astonishingly – it has been dubbed an "astonishing hypothesis"). Joseph Priestley clearly explicated this premise when he stated that "the powers of sensation or perception, and thought, as belonging to man, have never been found but in conjunction with a certain organized system of matter; and therefore, that those powers necessarily exist in, and depend upon, such a system. This, at least, must be our conclusion till it can be shown that these powers are incompatible with other known properties of the same substance; and for this I see no sort of pretence." (Priestley, 1777, p. 26) The *London Encyclopedia* of 1829, subtitled *Popular View of the Present State of Knowledge*, annotated (p. 637): "Dr. Priestley apprehends that sensation and thought necessarily result from the organization of the brain . . . but he professes to have no idea at all of the manner in which the power of perception results from organization and life."[2] Generally speaking, this 200-year-old account from the *Popular view of the Present State of Knowledge* very well describes our present-day theoretical understanding of the relationships between perceptions, or other mental activities, and features of the brain (despite the fact that the sheer amount of isolated data might seduce us to come to a more optimistic assessment).

For over 200 years, the premise that mental processes must be considered a function of the brain has been more or less commonplace. This has deluded us into overlooking the fact that, despite all the impressive insight we have gained into specific aspects, our theoretical understanding is next to nil of what exactly – within a specific range of mental phenomena and achievements – this function might actually be taken to be[3]. Neither are we currently capable of giving an adequate theoretical account of the fundamental mental processes in question, nor of producing a theoretical conception answering the question on which physical level of the brain's structure we might find the relevant principles that explain mental phenomena. Take, for example, perception, and how it is connected to functional organization of the brain: From a psychological point of view, we are not even close to having a theoretically adequate conception of perception; instead, we are still trapped in (Aristotelian) picture and similarity conceptions of perception (despite the fact that their inadequacy has also been obvious since the 17th century). From a biological point of view, we do not know on which physical level of the brain's organization the relevant principles are to be located that generate from physico-geometrical inputs "objects of perception" – be

[2]Priestley did not know, as little as we do today, *which* physical principles form the basis of specific mental processes, or *how* they do so (cf. Chomsky, 2010). Regardless of how closely we might scrutinize the brain – be it through a microscope, modern imaging devices, or future devices of yet greater precision –, we will always encounter physical objects and their given properties: neurons and synapses, neurotransmitters, ions, electrons and protons. Leibniz's mill-argument remains as valid as it was in the 17th century. William James cited a remark made by the physicist Tyndall in 1871, "in that lucky paragraph which has been quoted so often that every one knows it by heart" (James, 1890/1983, p. 150): "Granted that a definite thought, and a definite molecular action in the brain occur simultaneously, we do not possess the intellectual organ, nor apparently any rudiment of the organ, which would enable us to pass by a process of reasoning from the one phenomenon to the other. They appear together but we do not know why."

[3]This is patently obvious when we look at achievements of relatively simple organisms such as ants, bees, or the nematode *C.elegans*. The phenomenon of ant-navigation, the waggle dance of the honeybee or the complex behavior of nematodes have produced a large number of correlations between neuronal signals and behavior – without even coming close to being able to actually provide explanations for this behavior. In the case of *C. elegans*, the complete knowledge of the components of its biological hardware would constitute a particularly favorable situation for understanding its complex behavior (such as chemotaxis, thermotaxis and thermo memory, or mechanosensory reactions). "C. elegans responds behaviorally to the presence or absence of food in a plethora of ways. . . .Surprisingly little progress has been made in understanding these responses" (Thomas and Lockery, 2000, p. 156). This theoretical humbleness markedly contrasts with the kind of claims that pervade cognitive neuroscience and which seem to presume that the theoretical challenge is less serious when we are dealing with 100 billion neurons instead of 302. For the favorable conditions of much simpler kinds of organisms, the programmatic claim of the cognitive neurosciences that a theoretical understanding of behavioral achievements crucially has to refer to the neural level apparently falls short of its promise. A claim that is applicable to bees or nematodes but grossly founders, will hardly look promising in the case of exceedingly more complex neural systems.

it the levels of neurons or complex, dynamic systems of neurons, sub-cellular structures – such as the interaction of proteins –, or the level of quantum-theoretical processes etc[4]. To be sure, the premise that psychological phenomena are achieved by a biological organ, namely the brain, is a reasonable one, and unrivaled within the context of the natural sciences. However, as long as our relevant theoretical knowledge in psychology and biology continues to be poor, this premise remains rather inconsequential[5]. The better our substantial theoretical understanding is, the better are our chances to refine our theories by biological insights and to integrate them into an overarching theoretical perspective (e.g., Jenkins, 2000; Berwick and Chomsky, 2011).

Undeniably, during past decades we have gained a large amount of detailed information on the correlations between mental phenomena and neurobiological processes, such as the activation of neurons or metabolic processes in certain areas of the brain. But it would be a grave misunderstanding to mistake these findings for an *explanation* of psychological phenomena. In fact, these findings even increase the explanatory gap, for now we do not only have to explain the psychological phenomena as such, but also *why* these phenomena relate to precisely this, that, or some other neurobiological process. Most notably, we still lack a satisfactory theory of the electro-physical activity of the brain. Therefore, we presently cannot explain *why* the results from studies based on single cell recordings, local field potentials, EEG, fMRI, PET, etc., look the way they do, and not otherwise. As long as we lack a deeper understanding of the physical mechanisms underlying mental phenomena and achievements, the correlations between these phenomena and neurobiological processes will be of theoretical avail primarily to neurophysiology rather than to psychology[6].

Hence, our primordial task in psychology (as well as in biology) is to first lay a suitable fundament for a serious theoretical understanding by which we can then go beyond the centuries-old scientific truism stating that psychological phenomena are grounded in mechanisms of the brain (cf. e.g., Yolton, 1983) and gain a deeper understanding of this relationship. Dogmatic pronouncements about what is to be regarded as the "true" level of explanation for psychological phenomena can hardly be expected to foster our theoretical understanding. Moreover, such claims are based on a profound misunderstanding about the methodological principles of the natural sciences. Given that we presently know next to nothing about the physical principles underlying mental phenomena and achievements, there is no reason to assign the level of neurons a privileged explanatory role[7].

The 19th century was already marked by a tendency to reduce fundamental theoretical questions of psychology to questions about an alleged material substratum of mental phenomena. Corresponding ideas expressed a metaphysical worldview – "the physical is presumed to be epistemologically and metaphysically in good order, and the mental is questionable" (Stoljar, 2006, p. 46) – that, given the theoretical context of the time, was not completely unreasonable. It was a result of the materialism which, in the 19th century, was commonly accepted and regarded as an expression of a general and unified natural science perspective. But Priestley was already aware of the fact that the Newtonian concept of matter was calling the classical mechanistic conception into question. A new way of defining the "physical" was undermining the foundation of traditional ways of formulating the relation between material substratum and mental capacities. This also meant that classical materialism had been surpassed (Lange, 1902; Koyré, 1939/1978). Physics had lost any intuitively comprehensible notions of matter. Instead of matter, one now had to refer to the "physical," that is, the entirety of forces and entities that populate the currently best explanatory accounts of physics. The concept of the "physical" (cf. Stoljar, 2010), however, no longer lends itself to reductionist approaches. Rather, in the history of physics, this concept has proven itself to be open and constantly in flux. It may comprise entities such as gravitational fields, fermions, superstrings, or who knows what else in the future. Hence, we presently still have an insufficient and continuously changing understanding of what we regard as "physical." And since we lack a clear-cut notion of the "physical," we lack a clear-cut notion of the "non-physical" as well[8]. To put it briefly: As opposed to what was the case in the 17th century, a reductionist stance that assigns the level of neurons a

[4]Whatever the character of the physical principles underlying mental achievements turns out to be, it is to be expected that these principles are exceedingly more abstract and lie on a much deeper level of physical organization of the brain than a neural one. Presently, there is not much that speaks in favor of the highly speculative claim that the relevant physical principles underlying mental phenomena can be found on the coarse physical level of neural organization.

[5]Gardner (1987, p. 286/287), in the early phase of the *cognitive neurosciences,* reminded us once more of this matter of course: "One cannot have an adequate theory about anything the brain does unless one also has an adequate theory about the activity itself. It is not possible to study perception – even in its most fine-grained forms – without a theory of perception. . . . From this perspective, it is not possible to enter into the nervous system as a disinterested observer who is simply chronicling the facts as many neurosciences assume they are doing."

[6]Even in neurophysiology, we do not understand, for example, which information is coded by which principle in the sequence of action potentials (Rieke et al., 1997), or which components of a living cell can perform which kind of abstract computations. Also, as Brezina (2010) noted, "the neuronal wiring diagram alone is not sufficient to specify, and permit us to understand, the computation that underlies behavior." If we look top-down from behavioral achievements to an allegedly underlying neural base, we presently do not understand how the relevant computational operations could be realized by the given biological substrate. As Gallistel remarked (1997, S. 77f.): „We clearly do not understand how the nervous system computes. We do not know what are the foundations for its ability to compute. We do not understand how it carries out the small set of arithmetic and logical operations that are fundamental to any computation, the operations that are part of the basic instruction set in any computer ever developed, including massively parallel computers and neural net computers. We do not, for example, understand how neurons multiply, add, and compare the values of variables." These deficits are profound and cannot simply be papered over by enthusiastic proclamations.

[7]Given the fact that our theoretical understanding in neurophysiology itself is subject to constant change, it would be hard to even formulate such a claim. To assign a privileged role for the explanation of mental phenomena to a certain level of analysis of *current* neurophysiology would make as much sense as it would have in the 19th century to claim that the physics of that time played a privileged role in the explanation of chemical phenomena. To postulate, by contrast, that neurophysiology will *ultimately* turn out to play a privileged role for the explanation of mental phenomena would hardly provide interesting constraints for current attempts to develop explanatory accounts in psychology, and hence not be of much relevance.

[8]Bertrand Russell and Noam Chomsky, in particular, have emphasized the inherent openness of the concept of the physical. Because of this openness, attempts to establish a physical foundation for mental concepts are pointless. In order to even formulate them, "we have to have a notion of physical entity; we don't. It is a mere stipulation to include gravitational attraction, fields, Kekulé's structural formulas,

privileged role in the explanation of mental phenomena does no longer testify to an uncompromising and thorough natural science approach to mental phenomena. Rather, such reductionist approaches seriously conflict with methodological principles that have proven theoretically fruitful in the development of the natural sciences[9].

Accordingly, reductionist views that assign the level of neurons a privileged role in the explanation of mental phenomena have always remained a fringe position in the history of psychology. But this changed with the rise of so-called *cognitive neurosciences* in the 1980s. In the course of the proclamation of a new, interdisciplinary scientific approach to mental phenomena a new attitude has emerged that stipulates on which level of analysis, as of now, mental phenomena have to be investigated. If mental phenomena were investigated by studying neural processes, one would be able, so the claim goes, to directly get to the explanatory roots of mental phenomena, and hence to render an independent, psychological level of analysis more or less obsolete. During the last decades, such a conception has turned into a kind of widely accepted tacit background assumption whose appropriateness is regarded as a matter of course. It has been accompanied by corresponding strands in the philosophy of mind, where such conceptions also found their most explicit expression. An instructive example for corresponding claims is provided by Paul Churchland, a most prominent proponent of a new "neurophilosophy:" "We are now in a position to explain how our vivid sensory experience arises in the sensory cortex of our brains: how the smell of baking bread, the sound of an oboe, the taste of a peach, and the color of a sunrise are all embodied in a vast chorus of neural activity. . . . More centrally, we can now understand how the infant brain slowly develops a framework of concepts with which to comprehend the world. And we can see how the matured brain deploys that framework almost instantaneously: to recognize similarities, to grasp analogies, and to anticipate both the immediate and the distant future." (Churchland, 1995, p. 3) Admittedly, Churchland's stunning claim is a particularly fatuous example of neuro-propaganda. But it nevertheless is not untypical of the vigorous rhetoric that has accompanied the cognitive neurosciences in the past two decades[10].

Psychology, understood as the endeavor to theoretically understand mental phenomena and achievements, can naturally be regarded as a branch of biology[11]. In biology, it is regarded as a matter of course that relevant phenomena can and have to be investigated on separate levels of analysis (cf. Tinbergen, 1963). Psychology then constitutes an abstract level of analysis that is as autonomous explanatorily as levels of analyzes that characterize e.g., evolutionary theory, ethology, ecology, molecular biology, or cell biology. Accordingly, psychology has many intriguing points of contact to other fields of biology. In corresponding relationships, psychology has been guided by its explanatory demands rather than by dogmatic principles. This changed in the 1980s, with the proclamation of the *cognitive neurosciences,* when large areas of psychology committed themselves again, in a historic regress, to philosophy by adopting a metaphysical position that is scientifically unmotivated and philosophically dubious, at best. This metaphysical position, which assigns the level of neurons a privileged role in the explanation of mental phenomena, is in conflict with core methodological principles of the natural sciences, and as a result has proven unfavorable for the development of adequate explanatory frameworks for mental phenomena and achievements. In the course of these developments, the long-standing and fruitful relationship of psychology with other fields of biology has become burdened again by an element of dogmatism.

How could foundational research in psychology be dominated for decades by such a dogmatic approach, and how can it continue to be? The reasons for this development can hardly be expected to lie in the dynamics of theory development. Rather, they must be sought outside science itself, in factors pertaining to the psychology of science or the sociology of science. The following sections will briefly address five of these aspects.

FIVE REASONS FOR THE CURRENT DOMINANCE OF NEURO-ORIENTED CONCEPTIONS IN PSYCHOLOGY

What is it that confers to neuro-oriented perspectives the kind of attractiveness that they obviously currently have? The principle reason can likely be found in the fact that such views blend in well with some of our ordinary intuitions about psychological phenomena and modes of explanation. The ideas that underlie the *cognitive neurosciences* receive their apparent plausibility because they are based on common-sense intuitions, which are tacitly transposed into a context that superficially appears as natural science.

curved space-time, quarks, superstrings, etc., but not the processes, events, entities, and so on postulated in the study of mental aspects of the world." (Chomsky, 1996, p. 44).

[9]Revealingly, the history of science shows that theoretical efforts in the natural sciences have not primarily been concerned with issues of reduction to more basic levels of analyzes. Rather, these efforts aim to develop abstract explanatory accounts for certain types the phenomena and to achieve a unification of principles on which different types of theories are based (often allied with a structural conception of explanation). Kitcher (1989, p. 448), for instance, spoke of an "outmoded concept of reduction, which is tied to an inadequate account of scientific theories." Accordingly, radical reductionist perspectives are on the very fringes in the history of the theoretical development of the natural sciences. Chemistry provides an instructive example. Its success precisely rests on the fact that chemistry has abstained from reductionist ties to the physics of that time (see e.g., Brock, 1992; Clericuzio, 2010).

[10]This rhetoric is reminiscent of the one that, in the 1960s, accompanied artificial intelligence approaches to mental achievements, such as problem solving, pattern recognition, or machine translation. For instance, Feigenbaum and Feldman, 1963, p. 205) annunciated that a rather simple program that answers questions, which were formulated in a drastically restricted vocabulary and syntax, constituted an "important initial step" toward the goal of "discovering the information processing structure underlying the act of 'comprehending' or the process of 'understanding'." The literature of that time teems with excessive claims of this kind. Now that they have disappeared into thin air, we can clearly recognize their rhetoric character. However, as this kind of rhetoric fulfills certain social functions in science, it is hardly surprising that it perpetually reappears in different guises. It can ubiquitously be detected in the cognitive neurosciences. By rhetorically extrapolating certain kinds of isolated findings or local successes with respect to theoretical peripheral issues to fundamental ones, the actual gap between what has been promised and what has been achieved is concealed and attention is distracted from theoretical issues of real intellectual significance.

[11]This appraisal of psychology as a branch of natural science has been expressed by Karl Bühler, among many others. Bühler emphasized, in his *Die Krise der Psychologie*, that "it must not be physics but rather biology which this research perspective has to follow and toward which it has to be oriented." (Bühler, 1927, p. 71)

ONE: RELIANCE ON ORDINARY CONCEPTIONS OF "EXPLANATION" (AT THE EXPENSE OF NATURAL SCIENCE CONCEPTIONS OF EXPLANATORY DEPTH AND WIDTH)

In our ordinary modes of thinking we regard a phenomenon of interest as explained when we manage to trace it to something familiar and known. Our ordinary notions of explanation favor explanations in terms of tangible and visible processes or things. In other words, our ordinary modes of thinking exhibit a built-in preference for concretistic explanations. Because of this, explanations of mental phenomena (such as language, pattern recognition, or depression) are considered intuitively more plausible when they refer to concrete, "real" objects such as brain areas, neurotransmitters or genes, rather than to abstract theoretical entities. The inclination to uncritically take an investigation of the neural substratum of mental phenomena "as validation or invalidation of our ordinary view of the world" is aptly called "neuro-realism" by Racine et al., 2005, p. 160). The apparent plausibility of the conceptions underlying the cognitive neurosciences derives from our natural inclination to reduce the highly abstract and barely comprehended principles underlying mental phenomena to their alleged tangible material substratum. Such an idea could have been defended within the mechanistic worldview of the 17th century. But after the classical conception of matter dissolved, we cannot even formulate a corresponding conception in a coherent way.

Because we are in the grip of our ordinary conception of explanation, we have severe difficulties to accept the fact that the notion of explanation in the natural sciences radically deviates from the one in everyday life. In the natural sciences, what is "real" is entirely determined by what is postulated by our best theoretical accounts. The entities to which we refer in corresponding explanations had often previously been completely unknown (e.g., atoms, neutrinos, gravitational fields). Accordingly, in the natural sciences, we explain something familiar, namely a phenomenon, by something previously unknown, namely abstract, theoretical entities (cf. Koyré, 1939/1978, p. 155). Therefore, it is quite possible, and in fact likely, that whatever we regard as a satisfactory explanation for a given class of phenomena in the natural sciences, be it physics, biology, or psychology, will not conform to the explanatory needs of our ordinary modes of thinking. By the same token, explanations that we find satisfactory in our everyday modes of thinking will, as a rule, be regarded as theoretically inadequate or irrelevant in the natural sciences.

Prevalent notions of explanation and understanding employed in the cognitive neurosciences differ markedly from the ones underlying the natural sciences. This is instructively revealed by an assessment by Gazzaniga (2006): "The single most important insight that the cognitive neurosciences can offer ethicists is in understanding how the brain forms beliefs. . . . Belief formation is one of the most important areas in which cognitive neuroscience needs to teach something to ethicists and to the world." As we presently have not even the glimmer of an idea of "how the brains form beliefs," the notion of understanding, on which Gazzaniga's assessment is based, obviously has no bearing with the notion of theoretical understanding that is at the core of the natural sciences.

There is another characteristic difference between our ordinary notion of explanation and the one underlying the natural sciences. Our ordinary notion of explanation is a *local* one and aims to relate a concrete phenomenon to a concrete cause. In contrast, the history of science supports the view that "scientific explanations do not confer intelligibility on individual phenomena by showing them to be somehow natural, necessary, familiar, or inevitable" Friedman (1974, p. 18). In the natural sciences, the concept of explanation does not refer to isolated phenomena and their relations, but to properties of theories. The adequacy and value of a theory manifests itself in its explanatory power. The distinctive features of a successful theory pertaining to a certain range of phenomena are its *explanatory depth* and *explanatory width*. Explanatory depth refers to the property that the theory does not remain at the surface level of common-sense notions, but features a sufficiently deep deductive structure that connects it to experimental consequences which, from an everyday point of view, one would not necessarily have expected. Explanatory width refers to the potential of a theory to subsume classes of phenomena under a common explanatory framework that appear rather dissimilar und unrelated from our everyday perspective[12]. The natural sciences are willing to introduce and employ whatever kind of entities as long as they strengthen the explanatory depth and width of a theory. In this process, it is immaterial whether the entities to which explanations refer are intuitively plausible and in accord with our ordinary ideas on the corresponding set of phenomena.

In its preoccupation with local correlations between mental phenomena, on the one hand, and processes that are concrete and tangible from an everyday point of view, on the other, the notion of explanation underlying the cognitive neurosciences is closely related to our ordinary notions of explanation. It is this alliance that conveys to the cognitive neurosciences their apparent plausibility (as also witnessed by the widespread public interest in this type of local, neural "explanations" for mental phenomena)[13].

The wave of widespread fascination that accompanies brain imaging studies can only be understood when we take the concretistic inclinations of our ordinary modes of thinking into account. The illusion of understanding, which these studies tend to create (see Weisberg et al., 2008), stands in a stark contrast to their actual explanatory avail, which often remains at the surface of serious theoretical understanding. Using fMRI-studies on pain as an example, Hardcastle and Stewart (2009) evaluate the actual explanatory gain of such studies. They conclude that "most of these studies are not telling scientists anything that they did not already know from traditional psychological and clinical investigations." (p. 191) The outcome in theoretical understanding of the psychological phenomena under scrutiny provided by such studies does not go beyond the psychological insights invested into the work: "Thus far, it appears that the imaging technology has not improved our theoretical understanding of cognition; it has merely given us vivid illustrations of the cognitive processes that psychology had already surmised were there." (p. 192)

[12]Theories that exhibit no substantial deductive depth and explanatory width – especially "local theories," theoretically isolated models, or "theories" that crucially rely on common-sense intuitions – could only be called theories in a degenerate sense.

[13]An especially instructive example is the widespread public attention that mirror neurons have received and the fanciful interpretations that have been sparked by the corresponding findings (cf. Borg, 2007).

TWO: A GENERAL FOCUS ON MENTAL PHENOMENA THAT ARE SINGLED OUT BY EVERYDAY INTUITIONS (AT THE EXPENSE OF THEORETICALLY RELEVANT PHENOMENA)

The cognitive neurosciences share another aspect with our common-sense conception of mental phenomena and their explanation, namely the selection of phenomena that are regarded as being in particular need of explanation. In our ordinary mode of thinking, those phenomena are taken to be in particular need of explanation that cannot be taken for granted, and are unexpected and salient[14]. Conversely, we tend to take precisely those mental aspects as a matter of course – and hence as not being in need of explanation – that pertain to the most fundamental principles of our mental make-up. In fact, it is a core feature of our mental make-up that these aspects will pass unnoticed exactly *because* of their foundational nature. Wittgenstein's remark that we are unable to notice what is always before our eyes also applies to the scientific study of the mind[15]. Accordingly, when we are searching for neural correlates by which we hope to gain a better understanding of mental phenomena, we are inclined to take for granted what actually is in need of an explanation and to focus on phenomena that remain at the surface of our ordinary conception of phenomena.

As is well know from the history of the natural sciences, common-sense taxonomies are an inapt guide for the endeavor of achieving, within the framework of the natural sciences, a theoretical understanding of the mind. In fact, already in the case of physics, the tension between our ordinary intuitions, on the one hand, and the methodological principles underlying the natural sciences is so huge that its entire history can be understood as an attempt to dispense with common-sense classifications of phenomena and to instead follow lines of theorizing that are traced out by the development of successful explanatory accounts. The problems resulting from conflicts with common-sense intuitions are hence the rule rather than the exception in natural sciences that have not yet reached a mature state of theory development. Accordingly, it is hardly surprising that when mental processes become the object of scientific enquiry, our ordinary intuitions will, as a rule, be an even stronger impediment for the development of explanatory theories.

The characteristic gap between our ordinary intuitions and the kind of ideas that prove to be fruitful for the cumulative development of explanatory theories confronts us with particular difficulties in psychology. For the types of classification of phenomena that promise to yield, by appropriate theoretical idealizations, explanatorily fruitful theories about some abstract principles underlying our mental achievements may appear rather unnatural from our everyday perspective. Conversely, ordinary categorizations of mental phenomena will quite likely turn out to be inadequate and infertile for the development of explanatorily successful theories about relevant underlying principles, and hence cannot be expected to survive enquire. It would therefore hardly be surprising if almost all truly interesting psychological phenomena of our ordinary life remained untouched by a natural science approach and the kind theoretical insights it could yield into the nature of the mind – a point clearly recognized and expressed by Helmholtz (1862/1896), and Chomsky (e.g. 1988 p. 158f; 2000, p. 22f).

The overwhelming impact of our ordinary conceptions about mental activity manifests itself in the *cognitive neurosciences* already in the selection of phenomena that are singled out as particularly amendable for an "explanation" in terms of neural processes. Notwithstanding the usage of most recent technological advances, the general approach of the cognitive neurosciences resembles more that of a naturalist than a natural science approach. Concordant with a naturalist's attitude, the cognitive neurosciences seem to be more dedicated to finding and describing the peculiarities and particularities of the relation of ordinary psychological phenomena to neural ones than to gaining a deeper theoretical understanding of general principles. The influence of common-sense intuitions is sometimes masked or buried by the technical language employed – e.g., in fields such as "attention," "perception" or "thinking" – and hence hard to identify. Mostly, however, the impact of common-sense conceptions is rather obvious. This is especially true with respect to studies that have received widespread public attention – e.g., the question whether "trust" is modulated by oxytocin or testosterone, whether a person's "willingness to cooperate" may be manipulated by transcranial magnetic stimulation, or when neural correlates of feelings of "romantic love" were searched for in the caudate nucleus. Such investigations, notwithstanding their potential relevance for neurophysiology, suggest that already by linking psychological phenomena to neural processes a deeper theoretical understanding had been gained of the psychological phenomena in question. In fact, however, such studies usually take their starting point in everyday phenomena and remain at the surface of these ordinary conceptions in their conclusions. Research perspectives that focus on neural correlates tend to underestimate or overlook the deep theoretical challenge with which attempts are confronted to better understand the abstract psychological principles on which the corresponding phenomena and achievements are based. Neuro-oriented perspectives rather dodge these theoretical problems by purporting that they deal with the "true" basis of mental phenomena, and that a focus on neural correlates could therefore serve as a kind of explanatory shortcut. Corresponding approaches can hardly be

[14]The reigning conceptions of perception, for example, accede to the ordinary classification of perceptual phenomena into "normal" and "illusory" ones, despite the fact that the notion of "perceptual illusions" is based on a conflation and mix-up of different architectural components (cf. Mausfeld, 2002). This confusion has been pointed out again and again for more than 200 years and has been regarded as a severe obstacle to the development of appropriate theory about the principles underlying perception. Helmholtz (1855, p. 100), for instances, emphasized: "The senses cannot deceive us, they work according to their established immutable laws and cannot do otherwise. It is us who are mistaken in our apprehension of the sensory perception."

[15]An instructive example for a fundamental theoretical concept whose explanatory importance is almost entirely concealed from our ordinary intuitions about perception is the concept of a "perceptual object." Also, almost all structural and architectural aspects of our mental architecture are invisible to us. An instructive example for a core structural property of the perceptual system whose consequences pervade all of our mental activity is our capacity for maintaining different – and sometimes even conflicting – mental perspective simultaneously with respect to the same input situation. This capacity for "multiperspectivity" must be grounded in specific architectural and computational properties of the perceptual systems and subsequent systems. Despite the fact that these corresponding phenomena and achievements point to a core feature of our mental make-up, they have received little attention in the psychology of perception and in cognitive science (Mausfeld, 2011).

expected to yield a serious cumulative theoretical understanding, because their psychological explanatory gain remains mostly at the surface of our ordinary conceptualizations. All the same, they currently thrive because the alliance of neurophysiological perspectives with common-sense modes of thinking greatly attracts public attention and boosts the public visibility of psychology.

Conversely, it is hardly surprising in face of the history of the natural sciences, that an approach that focuses on a deeper theoretical understanding of single isolated (and idealized) "instruments" from the entire orchestra of our mental abilities radically conflicts with our ordinary intuitions and hence will hardly elicit widespread public attention. The kind of phenomena that we regard as particularly interesting from our ordinary perspective almost always pertains to achievements brought forth by interactions of a great many of these instruments. We therefore regard, in our ordinary modes of thinking, the person as a whole as a natural unit of analysis rather than isolated and idealized subsystems of the mind/brain. Accordingly, enquiries that focus on single specific subsystems that are amendable to theoretical idealizations must inevitably appear misguided and absurd. In contrast, theoretical notions that exhibit a close tie to our ordinary conceptions obviously receive a higher degree of plausibility and hence can be more easily communicated to a wider public.

The inclination to adhere to common-sense classifications of phenomena also is sustained by current science policies, which exhibit a strong tendency to favor interdisciplinary research projects. Of course, interdisciplinary projects can be advantageous in technologically oriented fields. In areas of foundational research, however, attempts to accelerate scientific progress by organizational policies that enforce or reward interdisciplinary work are likely to be pernicious, because interdisciplinary research perspectives increase the predisposition to remain, as a common theoretical denominator, close to the surface of common-sense intuitions. Hence, science policies that favor interdisciplinarity in fact tend to be an impediment to the advance of serious theoretical understanding – a clear lesson from the history of sciences that have achieved substantial theoretical content.

The closer a phenomenon remains to the surface of our everyday conceptualization, the easier it is to discover neural correlates and manipulate these correlations. These correlates – and the ability to systematically manipulate psychological phenomena – convey to us the impression that we have achieved a better understanding of the phenomena in questions. In line with our ordinary modes of thinking, we do not take these correlations as findings that themselves are in need of explanation but rather are inclined to regard them as explanations of the phenomena in question. However, the ability to successfully predict and manipulate a class of phenomena is, as the history of sciences amply illustrates, for deeper principle reasons entirely independent of their theoretical understanding (notwithstanding the fact that both aspects are usually correlated in mature sciences)[16]. With respect to our ordinary classifications of mental phenomena, one can, of course, identify a broad range of condition and effect variables – especially at the architectural periphery of performance aspects – and sometimes even be able to systematically manipulate these phenomena (which may be very useful for pragmatic purposes). If the phenomenon under scrutiny is appropriately chosen, a closer analysis of these condition and effect variables may in fact significantly contribute to a deeper theoretical understanding. But studies of this kind can in no way serve as a kind of surrogate for independent theoretical enquiries.

The inclination to draw on ordinary taxonomies of phenomena and to single out mental phenomena by everyday intuitions rather than by theoretical considerations brings forth especially detrimental effects on theoretical progress in foundational psychological research. Areas of the cognitive neurosciences that primarily pursue neurophysiological goals, or are based on animal studies are naturally much less prone to fall prey to corresponding psychological common-sense intuitions. In fact, ethological modes of thinking can provide an effective antidote against the influence of some types of unwarranted common-sense intuitions in foundational psychological research.

There are other types of tacit common-sense intuitions that can, in the context of the natural sciences, exert a detrimental influence on the development of theoretical understanding, viz. ones relating to methodological conceptions of how theoretical knowledge can be achieved. Whereas the deleterious influences discussed so far are specific to the domain of foundational psychological research, the adverse influence of common-sense intuitions pertaining to research methodology can affect all areas of the cognitive neurosciences (as well as other domains of the natural sciences). The next two sections deal with two types of common-sense intuitions that are related to methodological issues.

THREE: INDUCTIVIST (MIS-)CONCEPTIONS OF SCIENTIFIC PROGRESS

It is well-known from the history of the natural sciences that there are persistent and deep-rooted common-sense intuitions

[16]In psychology as well, applied areas and foundational enquiries follow independent paths of development. The institutional organization of a discipline must therefore take this difference – which is a categorical, not merely a gradual difference – into account. In his work on the history of science, Kuhn pointed out that almost no historical society has managed successfully to simultaneously advance technology *and* science. Ancient Greece, for example, favored science, *epistêmê*, over technology, *technê*, which was regarded as a finished heritage. Rome, in contrast, became famous for its technological achievements, but produced no notable science. "Germany during the century before World War II is the only nation that has managed simultaneously to support first-rate traditions in both science and technology. Institutional separation – the universities for *Wissenschaft* and the Technische Hochschulen for industry and crafts – is a likely cause of that unique success. As a first approximation, the historian of socioeconomic development would do well to treat science and technology as radically distinct enterprises, not unlike sciences and the arts." (Kuhn, 1971, p. 285) The widespread and persistent idea that science and technology developed hand in hand, and that science emerged as a kind of abstract continuation of technology is based on a profoundly distorted picture of their actual historical developments (see inter alia Koyré, 1956; Westfall, 1983; Gaukroger, 2006). The deepest distinguishing mark is the different relation to common-sense conceptions. While technology is predominantly motivated by common-sense conceptions and develops, conceptually and methodologically, in line with them, the major developments in science have been achieved by radically dispensing with common-sense intuitions. Science and technology developed, up to the 19th century, within independent conceptual frameworks, pursued different goals and were guided by different methodological principles. In terms of cognitive science, it seems reasonable to conjecture that science and technology exploit different mental capacities.

that influence in a deleterious way conceptions of how theoretical progress can be achieved. In our everyday conceptions, we think of knowledge as cumulating through the gathering of facts. The cognitive neurosciences seem to favor a corresponding inductivist conception of theory development, according to which explanatorily successful theories are almost automatically brought forth by a pertinacious accumulation of empirical findings, without the need for independent theoretical efforts. The sheer quantity of isolated data and experimental findings on the "neural foundations" of mental phenomena has conveyed an illusion of deeper understanding and has fostered the expectation that experimental investigations of a broad spectrum of neural variables will by themselves eventually yield a deeper theoretical understanding of the principles underlying mental phenomena and achievements. The corresponding a-theoretical and extremely inductivist conception characterizes large areas of the cognitive neurosciences. To mention just one example, such a conception seems to be explicitly endorsed by Yarkoni et al. (2010), who programmatically regarded it as the essential basis for "a cumulative science of human brain function." On their view, cumulative progress in the cognitive neurosciences has to be based "on accumulated evidence from hundreds or thousands of individual experiments" and therefore can only be ensured by "the synthesis and modeling of existing data, in addition to the collection of new data." The modern variants of Baconian inductivism have bred what Brenner (1997, p. 37), with respect to biology, called "the orgy of fact extraction in which everybody is currently engaged."

It is well known that corresponding inductivist conceptions of theory development profoundly misrepresent actual theory development in the history of the natural sciences. The quintessence of the insights into the character of modern (i.e., post-Galilean) science that have been achieved by corresponding enquires has been succinctly expressed by Koyré (1968, p. 18): "... science is primarily theory and not the gathering of 'facts'." The success of the natural sciences depends, more than on anything else, on achievements that pertain to explanatorily successful re-conceptualizations of previous modes of thinking. These re-conceptualizations in the theoretical frameworks employed for singling out relevant phenomena and for formulating theoretical principles cannot be derived from whatsoever type of "facts." Furthermore, and more importantly, inductivist conceptions of theory development are deeply flawed for reasons of principle. Because of this, the widespread idea is misguided that by "simply chronicling the facts" a theoretical understanding of mental phenomena could be attained, "as many neurosciences assume they are doing" (Gardner, 1987, p. 287).

Nevertheless, inductivist conceptions are a core element of the research methodology that tacitly governs large areas of the cognitive neurosciences and has turned the field into an extremely data-driven enterprise. Inductivist conceptions, furthermore, have given rise to a degenerate and "localist" notion of theory and have made experiments on isolated local issues almost an end in itself. As a result, the gap has widened between the abundance of isolated findings in well-established experimental paradigms, on the one hand, and the prospects for an explanatorily fruitful theoretical integration of these findings, on the other. Wide areas of the cognitive neurosciences exhibit a disdain for theoretical enquiries that is unprecedented in the entire early history of cognitive science, ranging from Alhazen to Helmholtz. A-theoretical and inductivist attitudes unavoidably go along with a negligence of long-term investments into cumulative theory-building. As a result, they rob and deplete, in the long run, the theoretical accomplishments that have been achieved by earlier generations. Also, a discipline that exhibits a disregard for independent theoretical efforts becomes susceptible to a *substitutive* form of theory development in which theoretical vogues follow another without any substantial cumulative theoretical progress.

FOUR: TAKING INVESTIGATIONS OF INDEPENDENT AND DEPEND VARIABLES AS AN END IN ITSELF

Fields that are marked by an inductivist orientation and accordingly attach little value to independent theoretical enquiries characteristically exhibit an overemphasis on investigations of "independent" and "dependent" variables which they almost regard as an end in itself. Again, this tendency is fostered and supported by our ordinary intuitions about mental processes. In our ordinary modes of thinking, we are mainly interested in the variables by which mental phenomena can be affected or manipulated ("Which factors bring forth a phenomenon or have an effect upon it?") and the corresponding effect variables ("What kind of observable effects are brought forth by different factors?"). In contrast, the internal principles and processes that are causally responsible for the kind of observed dependencies usually do not figure in our ordinary accounts.

Since these internal principles and mechanisms are practically invisible in our experience, we tend to focus on input variables, i.e., variables by which a phenomenon can be brought forth or modulated, and on output variables, i.e., variables that reflect the effects of the variations of the input variables on a phenomenon in question. At the same time, we ignore or gravely underestimate the complexity of the relevant underlying internal principles. This propensity is itself the result of our mental make-up (and hence a due research target for cognitive science). However, if this propensity is made into a methodological guideline for research, as in behaviorist paradigms, it clashes with basic methodological principles that have proven fruitful in the history of the natural sciences, and becomes pernicious to theory development. Still, due to the deeply entrenched impact of our ordinary modes of thinking, approaches that take investigations of independent and dependent variables (or input and output variables) as an end in itself pervade, often in a concealed way, large areas of psychology and cognitive science.

Corresponding perspectives are a heritage of behaviorism and have survived because they conform to our ordinary modes of thinking. Actually, the most crucial aspects of what has been referred to as the cognitive revolution have left no traces in the cognitive neurosciences. The so-called *cognitive turn* was, contrary to the intentions of its most prominent protagonists, more a change in terminology than a theoretical paradigm-shift. Behaviorist convictions have survived under the surface of new jargon. Two decades after the *cognitive turn* began, Miller (1979) observed:

"What seems to have happened is that many experimental psychologists who were studying human learning, perception, or thinking began to call themselves cognitive psychologists without changing in any obvious way what they had always been thinking and – as if they suddenly discovered they had been speaking cognitive psychology all their lives. So our victory may have been more modest than the written record would have led you to believe." What Miller regarded as a deplorable situation is proclaimed as a virtue by Roediger (2004), who hailed the fact that "behaviorism is alive and well and nothing 'has happened' to it. . . . Behaviorism is alive and most of us are behaviorists." Roediger's assessment certainly applies to cognitive neuroscience as well. Even if most cognitive neuroscientists might remain aloof to explicit endorsements of behaviorism, the shared methodological and meta-theoretical presumptions can be easily identified. While the idea of classical behaviorism that statements about mental states can be reduced to statements about behavior will hardly have any adherents anymore, cognitive neuroscientists nevertheless are prone to accept the idea that reference to behavior enters essentially into any adequate theoretical account of mental phenomena and achievements (cf. Strawson, 1994, p. 31ff). Fodor (2003) is thus right in stating: "In fact, though, practically all experimental psychologists and philosophers of mind continue to be behaviorists of one kind or other. They have just ceased to notice that they are." The disdain for independent theoretical enquiries and the preoccupation with investigating and modeling of input–output relations is a characteristic mark of the influence of both behavioristic thinking and common-sense conceptions of mental phenomena. The cognitive neurosciences have inherited from behaviorism what Mandler (Baars, 1986, p. 255) described as "fear of theory" and as an "antitheoretical point of view" that "characterized American Psychology and American science through the 19th century and into the 20th century."

The objection that in the natural sciences, be it physics, biology or psychology, input and output variables are the only empirically determinable and measurable data, is obviously true but pointless. Nothing follows from this fact that would allocate independent theoretical efforts a subordinate role. Measurement variables are an indispensable element of theory development but neither an end in itself nor the subject matter of theories. Just as physical theory does not aim at modeling measurement variables but rather at a *theoretical understanding* of a range of significant physical phenomena, psychology also aims at a theoretical understanding of specific classes of mental phenomena. The same is true of biology, where it would absurd to dodge the task of developing appropriate theories about the achievements of specific biological systems – be it bird vocalization, navigation of desert ants, or echolocation of bats – by an analysis of condition and effect variables. It is a profound misconception of the methodological principles of the natural sciences to favor investigations of input–output relations at the expense of theoretical enquiries. For instance, in solar astrophysics, it would be non-sensical to restrict enquiries to observational data, such as protuberances and electromagnetic energy emitted from the solar periphery. Rather, enquiries aim at a theoretical understanding of thermonuclear processes that are happening in the interior of the sun. Observable data are not an end in itself but an important ingredient for achieving a theoretical understanding of the corresponding phenomena. The same holds true for enquiries into the principles underlying mental phenomena and achievements (cf. Chomsky, 1980, p. 189). Behavior only provides a type of data and cannot be assigned a privileged status or an epistemological priority to other kinds of data[17].

Nevertheless, in the history of psychology, the analysis of input- and output-variables has always received more attention than work on the theoretical conceptualization of relevant internal principles[18]. This imbalance may be viewed as a kind of compensatory effort aiming to divert attention from theoretical incertitude and precariousness to the supposedly solid ground of methodologically highly sophisticated analyzes of condition and effect variables. While, for instance, perceptual psychology offers a rich variety of detailled research on the nature of stimulus conditions (including important mathematical analyzes in ecological physics), efforts to gain a theoretical understanding of internal principles are rather feeble, if not disdained. Needless to say, investigations of relevant stimulus conditions are an essential element in the process of developing appropriate theoretical accounts of internal principles. However, they cannot be used as a surrogate for independent theoretical efforts. Yet, there is a strong tendency in perceptual psychology to trivialize the problem of the internal principles by misconstruing internal structure as a kind of mirror-image of external world properties (in line with the naïve realism of our ordinary conception of perception). Because of this tendency, the deepest and most important insights into fundamental principles of perception, as expressed in the notion of "Gestalt" or in Helmholtz's sign conception of perception, remain theoretically unredeemed, and detached.

While investigations of the relation between input and output variables are an indispensable tool for developing successful explanatory theories, they are a methodological surface characteristics rather than a constitutive feature of a natural science approach. This to some extent also holds true for methodological features such as quantification or experimentation. Despite widespread misconceptions in this respect, a scientific approach does not become a natural science approach merely because it relies on these methods. A mimicking of surface characteristics of

[17] In fact, the integration of our cognitive functions into a behavioral expression, i.e., our actions, will probably be for principled reasons not amendable to the methodological principles of the natural sciences. In contrast to simpler organisms, whose behavior is bound by the effects of environmental stimulus variables, human behavior is brought forth by the entire orchestra of mental faculties, including those that exhibit features of spontaneity and creativity. This makes the prospects daunting if not hopeless to identify abstract explanatory principles of how input conditions, given an internal state, yield behavior. Accordingly, within the framework of the natural sciences, human behavior might remain an epistemological mystery. In contrast, theoretical enquiries into the nature of our conceptual endowment appear to be amendable to the kind of idealizations and abstractions that characterize the methodological principles underlying the explanatory success of the natural sciences.

[18] For similar reasons, aspects pertaining to statistical methods for evaluating evidence characteristically have received far more attention in psychology than aspects pertaining to the development of explanatorily successful re-conceptualizations. As Suppes (1962, p. 260) correctly noted: „It is a paradox of scientific method that the branches of empirical science that have the least theoretical developments often have the most sophisticated methods of evaluating evidence."

the natural sciences in the context of common-sense classifications of phenomena will inevitably bring about a cargo-cult science.

The tendency to illegitimately transfer to a science context tacit common-sense assumptions about the acquisition of theoretical knowledge is, needless to say, not a distinctive mark of the cognitive neurosciences but can be recognized in all fields of the natural sciences. Likewise, the problems that originate from the resulting a-theoretical attitudes have gained influence in all domains in the wake of the current academic *Zeitgeist*. However, the cognitive neurosciences, and psychology in general, are particularly in jeopardy because the detrimental effects of corresponding attitudes on theory development are much severer in sciences that have not yet reached a mature state of theory development.

FIVE: "SIMPLE" AND "OBJECTIVE" ACADEMIC PERFORMANCE INDICATORS (ABETTING THE EXPERIMENTAL STUDY OF INPUT–OUTPUT VARIABLES TO THE DISADVANTAGE OF CONTRIBUTIONS TO CUMULATIVE THEORY DEVELOPMENT)

In face of the plenitude and depth of theoretical insights and intuitions that have been achieved in the intellectual history of psychology, the currently predominant disdain for theoretical efforts is quite startling. The methodological strictures that pervade current cognitive neuroscience have yielded a degeneration of theoretical discourse that arguably is unprecedented in the field, maybe with the exception of the era of behaviorism. Psychological factors pertaining to the influence of common-sense conceptions will hardly suffice to explain this development. Rather, in the natural sciences, a large-scale disregard for theoretical efforts can only occur under favorable sociological conditions regarding the organization of science.

In past decades, the higher education sector has become increasingly been exposed to market forces and corporate organizational thinking, a process that is expedited by a burgeoning for-profit university industry. In the course of these developments, the mechanisms regulating individual academic careers have also been more and more subjected to criteria derived from economics[19]. Before these developments, the chief regulating value for academic careers had been "reputation." This value eludes a simple definition and hence cannot be captured by a set of simple indices. It presupposes an established and well-functioning system of intellectual values and standards and can only serve the function it is supposed to fulfill as long as the system of intellectual values is intact. The more established academic value systems erode, the stronger becomes the need to find a kind of simple and objective surrogate for "reputation." Economic modes of thinking promote the acceptance of superficially rational indices by which "reputation" as a regulating value can be superseded. The guiding idea in this process is that the actual yield of research investments manifests itself in an objective and straightforward way in a kind of symbolical capital, as expressed by indicators such as publication and citation indices, or other measures of "visibility." This type of symbolical capital again serves to accumulate further material capital, i.e., funding for people and apparatus, etc., which, in turn, is used to foster and advance the further accumulation of symbolical capital.

Accordingly, "external success" as indicated by "objective" sociological variables is presumed to mirror to some relevant extent "internal success," i.e., achievements that contribute to long-term theoretical progress. Such a presumption may be granted, under favorable circumstances, some validity in a mature science, i.e., a natural science that has dispensed with common-sense notions of phenomena and explanation. However, in light of the history of science, such a presumption is hardly warrantable in sciences such as psychology or cognitive neuroscience that are still in their earlier stages of theoretical development. In these sciences, institutional reliance on speciously "objective" variables of "external success" as key evaluation or control variables will inevitably incite the temptations of individual scientists to surreptitiously increase the values of their corresponding indices by suitable strategies. Ironically, these indices establish a kind of "egalitarian" element into academic careers, because they allow nearly anybody of enough determination to produce favorable indicator values in a virtually algorithmic fashion. Hence, it is hardly surprising that in the individual planning of academic careers, publication strategies are absorbing a rapidly growing share of productive energy.

The more "straightforward" and "objective" key control variables for academic careers are, the more vulnerable they become to individual attempts to optimize them in a direct way. Hence, these control variables become more and more useless as indicators of scientific achievements[20]. These dynamic effects, which are an inescapable and pernicious consequence of the current trend toward subjecting foundational research to economic criteria and imposing control and evaluation parameters that are alien to natural science, are conspicuous and well-known. Still, the currently predominant institutional organization of science proceeds on the presumption that the apparent face validity of such control variables and their acceptance by the general public outweigh their adverse consequences. It thereby dodges the demanding task of establishing regulation variables that are based on a fertile balance between criteria for "internal success" in terms of long-term

[19]The current dominance of the cognitive neurosciences cannot be understood without taking into account the predominant cultural and social values and the social context, in which these ideas have emerged. Mandler's (2007) observations, concerning the emergence of behaviorism, apply mutatis mutandis to cognitive neuroscience: "In science and philosophy, the period was marked by a pragmatic, atheoretical preoccupation with making things work – a trend to find its expression in psychology in functionalism and behaviorism. . . [Behaviorism] was consistent with a number of old and new American cultural and social value. It also had its kindred movements in such developments as the drive for scientific management and the time-and-motion studies of Frederick W. Taylor designed to make the American workers more productive at less cost." (Mandler, 2007, p. 49/111) As to other political factors, beyond economistic modes of thinking, that are currently shaping the development of the cognitive neurosciences, see for example the corresponding reports of the National Research Council (NRC) (2008, 2009)) and of the Royal Society (Flower, 2012)

[20]It is quite surprising that a discipline which rightfully prides itself of its high methodological standards has become addicted to the idea that a vital aspect of its internal organization almost exclusively can be based on indices whose validity and whose vulnerability to a direct surreptitious optimization have, up to now, not been systematically studied. The widespread acceptance of such indices derives from little more than their apparent plausibility with respect to our everyday conception of science – a criterion which we would hardly accept in situations of far lesser importance.

theoretical progress, and "external success" in terms of short-term sociological variables. As the institutional reliance on short-term indicators of "external success" continues to increase, a gap begins to open between the "yield" as expressed by these indicators, and actual scientific achievements in terms of theoretical progress. This is a process which, in the long run, will erode the individual and societal willingness to make long-term investments in theory development, on which the natural sciences vitally depend, and thus lead to a depletion of the theoretical substratum that has been achieved by previous generations.

On an individual level, the currently effective academic control and evaluation parameters have fuelled tendencies to focus on problems which will quickly and predictably transform into publishable results. Patently, the safest way to swiftly ensure auspicious values on relevant evaluation parameters is a trouble-free production, in an assembly line manner, of more or less theoretically trifling or isolated experimental results. This state of affairs has in turn affected and shaped the mode of operation in research routine. The cognitive neurosciences appear to be particularly vulnerable to such deleterious developments. Once one has mastered apparatus and jargon of the neurosciences, one can deal with the entire spectrum of mental phenomena that fascinate us in our ordinary life, ranging from perception, thinking, or decision-making to depression, cooperative behavior, political attitudes, advertising, or religion. The procedural schema of corresponding neuro-oriented research does rarely require to work ones way into specific and sophisticated psychological theories and to come to grips with specific and often deep and long-standing theoretical issues that are inescapably attached to psychological achievements. Accordingly, the cognitive neurosciences offer particularly fortunate conditions for attempts to swiftly optimize the criteria that are currently regarded as control variables for success in the academic market. Corresponding individual efforts perforce will yield a growing percentage of experimental studies that take investigations of theoretically more or less superficial independent and dependent variables, within received paradigms, as an end in itself. While the presently prevailing emphasis on rather simple quantitative indices of "visibility" and short-term achievements is itself the offspring of the predominating inductivist and data-oriented (mis-)conception of science, it will in turn promote and amplify the proliferation of a-theoretical attitudes in the empirical sciences.

The a-theoretical attitude presently prevailing in the cognitive neurosciences can accordingly be attributed to the effects of two interlinked types of factors, both of which are alien to natural science: psychological factors related to common-sense conceptions regarding the classification of phenomena and notions of explanation, on the one hand, and sociological factors related to economically oriented control and evaluation variables, on the other. Of course, not only cognitive neuroscience but psychology as a whole has been affected by these developments, in the course of which guiding ideals and methodological principles of what is called modern science are at risk of sinking into oblivion.

These more cursory notes on some unwarranted tacit assumptions in cognitive neuroscience have been occasioned by the observation of a discrepancy between the promises and the extraordinary public and institutional success of the cognitive neurosciences, on the one hand, and a less favorable assessment of actual achievements in our theoretical understanding of mental phenomena, on the other. The endeavor to gain a better theoretical understanding, within the framework of the natural sciences, of the principles underlying mental phenomena is undoubtedly of paramount importance, not only for psychology but for the program of the natural sciences in general. Consequently, it is of some significance to identify factors that presumably are adversary to this endeavor.

REFERENCES

Baars, B. J. (1986). *The Cognitive Revolution in Psychology*. New York: Guilford Press.

Barlow, H. B. (1983). "Understanding natural vision," in *Physical and Biological Processing of Images*, eds O. J. Braddick and A. C. Sleigh, (Berlin: Springer), 2–14.

Berwick, R. C., and Chomsky, N. (2011). "The Biolinguistic Program: The current state of its evolution and development," in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 19–41.

Borg, E. (2007). *If Mirror Neurons are the Answer, What was the Question? J. Conscious. Stud.* 14, 5–19.

Brenner, S. (1997). Loose ends. *Curr. Biol.* 7, R202.

Brezina, V. (2010). Beyond the wiring diagram: signalling through complex neuromodulator networks. *Philos. Trans. R. Soc. Lond. B* 365, 2363–2374.

Brock, W. H. (1992). *The Fontana History of Chemistry*. London: Fontana Press.

Bühler, K. (1927). *Die Krise der Psychologie*. Jena: Fischer.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1988). *Language and Problems of Knowledge. The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1996). *Powers and Prospects. Reflections on Human Nature and the Social Order*. London: Pluto Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2010). "The mysteries of nature. How deeply hidden?," in *Chomsky Notebook*, eds J. Bricmont and J. Franck (New York: Columbia University Press), 1–33.

Churchland, P. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.

Clericuzio, A. (2010). "Sooty Empiricks" and Natural Philosophers: The status of chemistry in the seventeenth century. *Sci. Context* 23, 329–350.

Feigenbaum, E. A., and Feldman, J. (eds) (1963). *Computers and Thought*. New York: McGraw-Hill.

Flower, R. et al. (2012). *Neuroscience, Conflict and Security*. Report of the Royal Society, RS Policy document 06/11. London: The Royal Society.

Fodor, J. A. (2003). Is it a bird? Problems with old and new approaches to the theory of concepts. *Times Literary Supplement*, January 17, 2003, 3–4.

Friedman, M. (1974). Explanation and scientific understanding. *J. Philos.* 71, 5–19.

Gallistel, R. (1997). "Neurons and memory," in *Conversations in the Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 71–89.

Gardner, H. (1987). *The Mind's New Science. A History of the Cognitive Revolution*. New York: Basic Books.

Gaukroger, S. (2006). *The Emergence of a Scientific Culture: Science and the Shaping of Modernity 1210-1685*. Oxford: Oxford University Press.

Gazzaniga, M. S. (2006). "Facts, fictions and the future of neuroethics," in *Neuroethics: Defining the Issues in Theory, Practice, and Policy*, ed. J. Illes (Oxford: Oxford University Press), 141–148.

Hardcastle, V. G., and Stewart, C.M. (2009). "fMRI: A modern cerebrascope? The case of pain," in *The Oxford Handbook of Philosophy and Neuroscience*, ed. J. Bickle (Oxford: Oxford University Press), 200–225.

Helmholtz, H. V. (1855). *Über das Sehen des Menschen. In Vorträge und Reden*, 4th Edn, Vol. 1, 1896. Braunschweig: Vieweg.

Helmholtz, H. (1862/1896). *Über das Verhältniss der Naturwissenschaften zur Gesammtheit der Wissenschaft. In Vorträge und Reden*, Vol. 1. Braunschweig: Vieweg.

James, W. (1890/1983). *The Principles of Psychology.* Cambridge, MA: Harvard University Press.

Jenkins, L. (2000). *Biolinguistics.* Cambridge: Cambridge University Press.

Kitcher, P. (1989). "Explanatory unification and the causal structure of the world," in *Scientific Explanation*, eds P. Kitcher and W. Salmon (Minneapolis: University of Minnesota), 410–505.

Koyré, A. (1939/1978). *Galileo Studies.* Hassocks: The Harvester Press.

Koyré, A. (1956). The origins of modern science: a new interpretation. *Diogenes* 4, 1–22.

Koyré, A. (1968). *Metaphysics and Measurement: Essays in Scientific Revolution.* London: Chapman and Hall.

Kuhn, T. S. (1971). The relations between history and history of science. *Daedalus* 100, 271–304.

Lange, F. A. (1902). *Geschichte des Materialismus*, Vol. 2. Leipzig: Baedeker.

Mausfeld, R. (2002). "The physicalistic trap in perception theory," in *Perception and the Physical World*, eds D. Heyer and R. Mausfeld (Chichester: Wiley), 75–112.

Mausfeld, R. (2010a). "The perception of material qualities and the internal semantics of the perceptual system," in *Perception Beyond Inference. The. (Information)Content Of Visual Processes*, eds L. Albertazzi, G. van Tonder, and D. Vishwanath (Cambridge, MA: MIT Press), 159–200.

Mausfeld, R. (2010b). Psychologie, Biologie, kognitive Neurowissenschaften. Zur gegenwärtigen Dominanz neuroreduktionistischer Positionen und zu ihren stillschweigenden Grundannahmen. *Psychol. Rundsch.* 61, 180–190.

Mausfeld, R. (2011). "Intrinsic multiperspectivity. Conceptual forms and the functional architecture of the perceptual system," in *Interdisciplinary Anthropology: Continuing Evolution of Man*, eds W. Welsch, W. J. Singer, and A. Wunder (Berlin: Springer), 19–54.

MacLeod, D. I. A. (2010). "Into the neural maze," in *Color Ontology and Color Science*, eds J. Cohen and M. Matthen (Cambridge, MA: MIT Press), 151–178.

Mandler, G. (2007). *A History of Modern Experimental Psychology: From James and Wundt to Cognitive Science.* Cambridge, MA: MIT Press.

Miller, G. A. (1979). *A Very Personal History (Occasional Paper No. 1).* Cambridge, MA: Massachusetts Institute of Technology, Center for Cognitive Science.

National Research Council (NRC). (2008). *Emerging Cognitive Neuroscience and Related Technologies.* Washington, D.C.: National Academics Press.

National Research Council (NRC). (2009). *Opportunities in Neuroscience for Future Army Applications.* Washington, D.C.: National Academies Press.

Priestley, J. (1777). *Disquisitions Relating to Matter and Spirit.* London: Johnson.

Racine, E., Bar-Ilan, O., and Illes, J. (2005). fMRI in the public eye. *Nat. Rev. Neurosci.* 6, 159–164.

Rieke F., Warland D., de Ruyter van Steveninck R., and Bialek W. (1997). *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press.

Roediger, H. L. (2004). What happened to behaviourism. *Association for Psychological Science, Presidential column, APS Observer*, 17(3).

Stoljar, D. (2006). *Ignorance and Imagination.* Oxford: Oxford University Press.

Stoljar, D. (2010). *Physicalism. London: Routledge. The London Encyclopaedia, Or, Universal Dictionary of Science, Art, Literature, and Practical Mechanics (1829)*, Vol. XX. London: Tegg.

Strawson, G. (1994). *Mental Reality.* Cambridge, MA: MIT Press.

Suppes, P. (1962). "Models of data," in *Logic, Methodology and Philosophy of Science*, eds E. Nagel, P. Suppes, and A. Tarski (Stanford: Stanford University Press), 252–261.

Thomas, J. H. and Lockery, S. R. (2000). "Neurobiology," in *C. elegans: A Practical Approach*, ed. I. A. Hope (Oxford: Oxford University Press), 143–180.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20, 410–433.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., and Gray, J. R. (2008). The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* 20, 470–477.

Westfall, R. S. (1983). "Robert Hooke, mechanical technology, and scientific investigation," in *The Uses of Science in the Age of Newton*, ed. J. G. Burke, (Los Angeles: University of California Press), 85–110.

Yarkoni, T., Poldrack, R. A., and Van Essen, D. C. (2010). Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends Cogn. Sci.* 14, 489–496.

Yolton, J. W. (1983). *Thinking Matter. Materialism in Eighteenth Century Britain.* Oxford: Blackwell.

Zeki, S. (1998). Art and the brain. *Daedalus* 127, 71–103.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 August 2011; paper pending published: 22 September 2011; accepted: 21 February 2012; published online: 14 March 2012.

*Citation: Mausfeld R (2012) On some unwarranted tacit assumptions in cognitive neuroscience. Front. Psychology **3**:67. doi: 10.3389/fpsyg.2012.00067*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.