**From Brain to Mind and Back with Neuroimaging: Searching for Suitable Inference Models**

*or, alternatively:*
*(I can hold the talk in either language, as recommended by the organizers)*

**Mit der Bildgebung vom Gehirn zum Geist und Zurück: Auf der Suche nach Geeigneten Inferenzmodellen**

*Stephan Schleim, M.A., Division of Medical Psychology, University Clinics Bonn*

More than 50 years have passed since the first recordings of cerebral blood flow in feline species, marking the birth of neuroimaging. After breakthroughs in the 1970s which allowed to measure changes in blood flow related to neural activity in human subjects using positron emission tomography (PET), the availability of functional magnetic resonance imaging (fMRI) since the 1990s, the "decade of the brain", has lead to an increasing number of neuroimaging studies investigating brain function in healthy subjects as well as in clinical conditions. PET and fMRI scanners, among other technologies, are now widespread and researchers from many different disciplines like psychology, psychiatry, linguistics, and even philosophy, to name just a few, are performing experiments in order to answer questions about the underlying neural conditions of various cognitive processes such as perception, imagery, mentalizing, and moral decision making.

However, one of the major limitations of such imaging methods is that their findings are correlative and do not allow inferences about causal relations in neither the brain, nor the mind. Furthermore, these haemodynamic correlations are not explanatory in themselves, but rather observations in need of further explanation. Even though the combination of different fields of investigation, for example, lesion studies, neural projections, and inhibitory approaches using transcranial magnetic stimulation or other methods, increases the available knowledge about the functioning of several brain regions, neuroimaging still relies mostly on one particular kind of inference model, namely, *reverse inferences* (Poldrack, 2006). That is, the findings of a given study are explained through the cognitive processes which allegedly caused or accompanied the brain activation of other neuroimaging studies. Yet, the same inferential limitations apply to these further studies invoked as explanation for the respective finding.

There are two major problems inherent in this approach: First, reverse inferences strongly rely on the degree of functional specification of brain regions – the more functions, the less plausible it is

*eo ipso* why one particular cognitive process should have led to the reported activation and not another; however, how specialized brain regions are is still a matter of controversy. Second, these inferences also rely on the level of specification of cognitive processes – the more general one is defined (e.g. "social cognition" as opposed to "mentalizing" or "moral reasoning" which both belong to the general field of the former), the less specific the findings are and the less inferential value reverse inferences do confer. Conceptual confusions have been documented and criticized before (Bennett & Hacker, 2003; Bennett, Dennett, Hacker & Searle, 2007) which suggests that there is no common standard in the field of neuroscience. Thus, before invoking other studies for reverse inferences, it is necessary to check conceptual congruence in each particular case. Radically speaking, the strong reliance on this inference model which explains findings through other findings themselves in need of explanation may ultimately leave neuroimaging findings unexplained, at least insofar as the former two limitations are not considered appropriately.

In my talk, I will start out with the explanation of reverse inferences, referring to examples from the area of social cognition. These are particularly interesting for philosophy, since they refer to capacities related to the idea of man and which are themselves objects of philosophical inquiry (e.g. moral reasoning). These examples will illustrate the limitations of such inferences. In the next step, I will present in how far advanced methods (e.g. so-called "mind reading", Schleim & Walter, 2007; Schleim, 2008) and evidence from neuroscience beyond neuroimaging can increase the explanatory value of inferences. My major conclusion will be that for the understanding of brain function and correlative findings a closer cooperation of cognitive scientists, psychologists, and even philosophers with neurosciencists is necessary, taking into account the temporal dynamics of cognitive processes as well as the methodological constraints on the different levels of neuroscientific investigation.

**References**

Bennett, M. R. & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Blackwell, Malden, MA.

Bennett, M. R., Dennett, D., Hacker, P. M. S. & Searle, J. (2007). *Neuroscience & Philosophy: Brain, Mind & Language*. Columbia University Press, New York, NY.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *TRENDS in Cognitive Sciences* 10: 59-63.

Schleim, S. (2008). *Gedankenlesen: Pionierarbeit der Hirnforschung*. Heise Verlag, Hannover,

Schleim, S. & Walter, H. (2007). Gedankenlesen mit dem Hirnscanner. *Nervenheilkunde* 26: 505-510.