

## ***Abduction, Reverse Inference, and the Neuroscience of Moral Reasoning***

Gustavo Cevolani (IMT School for Advanced Studies, Lucca) and Luca Cecchetti (IMT School for Advanced Studies, Lucca)

Philosophical debates on the neural basis of moral reasoning and cognition mostly revolve around the “normativity problem”: what, if anything, empirical findings in the context of experimental tasks tell us about the nature and validity of moral norms and (meta-)ethical theories (e.g., Greene 2008, Berker 2009, Roskies 2016).

In this talk, we argue that a different problem is likely in need of more urgent attention. This is what we call the “inference problem”: even assuming that neuroscientific findings are indeed relevant for ethical theorizing, are these results robust enough to sustain the normative implications we want to derive from them? The answer hinges on the reliability of a widespread inferential practice, leading from the “low” level of patterns of neural activation to the “high” level of cognitive processes. The weakness of this “reverse inference” strategy has recently attracted a great deal of attention from both scientists and epistemologists, and its present methodological status is highly controversial (e.g., Poldrack 2006, 2011; Machery 2014; Glymour and Hanson 2016; Nathan and Del Pinal 2017).

Starting from some classical work in the philosophy of science, we explore the implications of the inference problem for current research in so-called neuroscience of ethics (Roskies 2016). In particular, we argue that Charles Sanders Peirce’s treatment of abductive reasoning, and the related notion of “inference to the best explanation”, can help in understanding and systematizing different attempts to make current inferential strategies more reliable (Schurz 2017). As a case study, we focus on the NeuroSynth project of Yarconi *et al.* (2011), which provides a large-scale synthesis of neuroimaging findings explicitly designed to address the reverse inference problem and related issues. We then argue for an approach which, clarifying the methodological issues underlying current scientific advances, can shed new light on ongoing debates at the interface of neuroscience and moral philosophy.

### **References**

- Berker, Selim. 2009. ‘The Normative Insignificance of Neuroscience’. *Philosophy & Public Affairs* 37 (4): 293–329.
- Glymour, Clark, and Catherine Hanson. 2016. ‘Reverse Inference in Neuropsychology’. *The British Journal for the Philosophy of Science* 67 (4): 1139–53.
- Greene, Joshua D. 2008. ‘The Secret Joke of Kant’s Soul’. *Moral Psychology* 3: 35–79.
- Machery, Edouard. 2014. ‘In Defense of Reverse Inference’. *The British Journal for the Philosophy of Science* 65 (2): 251–67.
- Nathan, Marco J., and Guillermo Del Pinal. 2017. ‘The Future of Cognitive Neuroscience? Reverse Inference in Focus’. *Philosophy Compass* 12 (7): <https://doi.org/10.1111/phc3.12427>.
- Poldrack, R. 2006. ‘Can Cognitive Processes Be Inferred from Neuroimaging Data?’ *Trends in Cognitive Sciences* 10 (2): 59–63.
- Poldrack, Russell A. 2011. ‘Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding?’. *Neuron* 72 (5): 692–97.
- Roskies, Adina. 2016. “Neuroethics.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2016. Metaphysics Research Lab, Stanford University.
- Schurz, Gerhard. 2017. “Patterns of Abductive Inference.” In *Springer Handbook of Model-Based Science*, edited by Lorenzo Magnani and Tommaso Bertolotti, 151–73.
- Yarconi, Tal, Russell A. Poldrack, Thomas E. Nichols, David C. Van Essen, and Tor D. Wager. 2011. ‘Large-Scale Automated Synthesis of Human Functional Neuroimaging Data’. *Nature Methods* 8 (8): 665–70.