Markus I. Eronen

Reduction in Philosophy of Mind

A Pluralistic Account

EPISTEMISCHE STUDIEN
Schriften zur Erkenntnis- und Wissenschaftstheorie

Herausgegeben von / Edited by

Michael Esfeld • Stephan Hartmann • Albert Newen

Band 24 / Volume 24

Markus I. Eronen

Reduction in Philosophy of Mind

A Pluralistic Account

ontos
verlag

Frankfurt I Paris I Lancaster I New Brunswick

Bibliographic information published by Deutsche Nationalbibliothek
The Deutsche Nastionalbibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

North and South America by
Transaction Books
Rutgers University
Piscataway, NJ 08854-8042
trans@transactionpub.com

United Kingdom, Ire, Iceland, Turkey, Malta, Portugal by
Gazelle Books Services Limited
White Cross Mills
Hightown
LANCASTER, LA1 4XS
sales@gazellebooks.co.uk

Livraison pour la France et la Belgique:
Librairie Philosophique J.Vrin
6, place de la Sorbonne ; F-75005 PARIS
Tel. +33 (0)1 43 54 03 47 ; Fax +33 (0)1 43 54 48 18
www.vrin.fr

P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 978-3-86838-125-2

2011

Printed on acid-free paper
ISO-Norm 970-6
FSC-certified (Forest Stewardship Council)
This hardcover binding meets the International Library standard

Printed in Germany
by CPI buch bücher.de

CONTENTS

Acknowledgements

The process that lead to this book started in 2003, when I began writing my master's thesis on emergence in philosophy of mind. My supervisor at the University of Helsinki, Sami Pihlström, introduced me to Achim Stephan's work on emergence. After I had finished my studies in Helsinki, I visited Achim in Osnabrück for the summer semester 2005, and he agreed to be my PhD supervisor. During that summer he also introduced me to Bob Richardson, who in turn introduced me to a whole new way of doing philosophy, and helped me see the shortcomings of traditional analytic philosophy of mind. That was when the basic idea of my PhD thesis began to take shape: to criticize reduction in philosophy of mind from the point of view of philosophy of science. In the end, Bob also became my second PhD supervisor. These three people (Sami, Achim, and Bob) have influenced my philosophical development probably more than they know, and I am in deep gratitude for them. Achim was also my main supervisor during the four years in Osnabrück, and I could not imagine a more helpful, supporting, and kind "Doktorvater."

In addition, I would like to thank the following people (in no particular order):

- Everybody at the Center for Research on Networked Learning and Knowledge Building (University of Helsinki), in particular Kai Hakkarainen, Liisa Ilomäki, and Sami Paavola, for providing an unusual but very inspiring start for my scientific career, and for letting me go to Osnabrück for the summer semester 2005, even though I was supposed to be performing my non-military service
- Vera Hoffmann-Kolss for helping develop my views on causation that play a crucial role in this book, for comments on several presentations and drafts related to this thesis, and for taking me in as a co-teacher for two inspiring seminars on philosophy of science
- Sven Walter for inviting me to be a co-author of an article on reduction that substantially influenced parts of this book (Walter &

Eronen 2011), and for kindly agreeing to be the third reviewer of my PhD thesis

- Rafael Hüntelmann at Ontos Verlag for helping to prepare this book for publication
- Brian McLaughlin for extending my philosophical understanding with the several seminars and talks given at the University of Osnabrück between 2006 and 2010, and for comments on earlier drafts of Chapter 8
- Ilaria Serafini for the friendship, for all the crazy ideas, and for making the years in Osnabrück surprisingly exciting and interesting
- Miriam Kyselo for countless philosophical discussions, an unforgettable thesis-writing trip to Morocco, and for making me smile and laugh, even when I tried to work
- The PhD program of the Institute of Cognitive Science, and especially its coordinators Carla Umbach and Peter Bosch, for all their help and support
- Other members of the PhD programme, particularly Rudi, Sascha, Hartmut, Katya, and Mikko, for all the interdisciplinary input
- Dan Brooks for the conversations over beer, for having the same philosophical vision as I do, and particularly for shedding light on the details of Woodward's theory
- Jani Raerinne for very detailed and constructively critical comments on several drafts of this book, since the very beginning
- Ulas Türkmen for the passionate philosophical dialogues (and monologues) and the technical support
- The "Mäyränkaatajat" group in Helsinki (Antti, Heikki, Ilkka, Jani, Janne, Jussi, Kari, Reino, Touko): these guys pull me down when I'm riding too high, and lift me up when I've fallen too deep
- My brother Jussi for all the practical and brotherly support and guidance
- My parents for always believing in me and supporting me in every possible way

- Kai Kuikkaniemi for the walks and talks that have helped me clarify my ideas through the years
- Petri Savolainen for long discussions on the nature of existence, science, and God that have also influenced the ideas in this book
- Sami Airaksinen, the second-greatest poet of Nilsiä, for being my friend in arts and for helping develop my writing skills through the years
- Antti-Jussi and Nina Pyykkönen, among other things for offering me a job as an *au pair* in case I ran out of funding
- Konstantin Todorov for being my best friend during the years in Osnabrück – keep with the snakes, stupid rocky!
- All the people who have assisted or supported me but who I forgot to mention here (sorry for that)

This book is a revised and updated version of my PhD thesis, which I defended in October 2010 at the University of Osnabrück. My PhD research was financially supported by generous grants from Helsingin Sanomain 100-vuotissäätiö (2006-2007), DAAD (2007-2009), and the Finnish Cultural Foundation (2009-2010). In addition, the PhD program of the University of Osnabrück provided funding for my numerous conference and workshop trips between 2006 and 2010. I am very grateful to these organizations for making this project possible.

While finishing this book, I was first a junior fellow at the Hanse-Wissenschaftskolleg in Delmenhorst and then a postdoctoral researcher at the Ruhr-Universität Bochum, in the group of Albert Newen. I thank these institutes for their support and for allowing me to invest time in preparing this book for publication.

Finally, I would like to thank Laura Bringmann for very useful comments on different versions of this book, most importantly the penultimate version, and for keeping me sane during the busy months of finishing my thesis and this book. In addition, thank you for making me happy. Nobody does it quite the way you do, nobody does it better.

Introduction

The idea of *reduction* has surfaced in different forms throughout the history of science and philosophy. Thales took water to be the fundamental principle of all things; Leucippus and Democritus argued that everything is composed of small, indivisible atoms; Galileo and Newton tried to explain all motion with a few basic laws; 17th century mechanism conceived of everything in terms of the motions and collisions of particles of matter; British Empiricism held that all knowledge is derived from experiential knowledge; current physicists are searching for the TOE, the "Theory Of Everything," that would unify the electromagnetic and the weak and strong nuclear forces with gravity. In a broad sense, all of these projects can be understood as (attempted) reductions, as they aim at revealing some kind of unity or simplicity behind the appearance of plurality or complexity. In philosophy of mind, reduction has figured prominently in the issue of the relation between the mind and the brain: Does the mind *reduce* to the brain? Do mental explanations *reduce* to neuroscientific explanations? Does psychology as a science *reduce* to neuroscience? And so on.

But what exactly is "reduction"? Traditionally, it has been understood as the derivation of a theory to be reduced from a more fundamental theory. However, it is now widely accepted in philosophy of science that this traditional view fails to characterize actual scientific practice, or actual relations between sciences, at least when it comes to psychology and neuroscience. In philosophy of mind, reduction is commonly conceived as "functional reduction," where reduction consists in defining a property[1] functionally and then finding the physical realizers that perform this function, but this model hardly fits scientific practice any better than the traditional model, and is plagued with philosophical problems.

[1] Often it would be more natural to talk of mental capacities or functions or processes, but following the venerable tradition in philosophy of mind, I mainly talk about mental "properties" in this book (without assuming any particular metaphysical theory of properties). In some contexts I use the term "state" instead of "property," but this subtle difference has no relevance for the arguments. I also talk about "mental" and "psychological" properties interchangeably and make no distinction between them.

In this book, I draw from recent developments in philosophy of science, and explore their consequences for the debates on reduction in philosophy of mind. I elaborate a pluralistic account of reduction, and show how and why more strongly reductionistic approaches fail. A pluralistic account of reduction might sound strange and contradictory. Aren't reduction and pluralism mutually exclusive? What I hope to show in this thesis is that the answer is no. The kind of pluralism defended here is compatible with certain kinds of reductions or reductive explanations. And I argue that, in fact, there are no reductions to be expected in any stronger sense.

This thesis is primarily intended as a contribution to the philosophy of mind and cognitive science, and what I am focusing on is the purported reduction of psychology (understood as an empirical science, not "folk psychology") to neuroscience.[2] The positions and arguments defended in this thesis do not necessarily apply to relations between other sciences, although I am happy if they do. The main target of my criticism is traditional analytic philosophy of mind, which has been largely guided by conceptual analysis and formal methods instead of actual science. If philosophy of mind is brought closer to actual science, it can also be more relevant to scientific endeavors of understanding the mind and consciousness.[3] One of the most prominent proponents of the traditional analytic philosophy of mind is Jaegwon Kim, who receives the most attention in this thesis, partly because I am more familiar with his work

[2] With "psychology" I mean the empirical study of human behavior and the mind, and with "neuroscience" the empirical study of the human nervous system. Of course, this distinction is becoming increasingly blurry, and is to some extent conventional. I make the distinction mainly for the sake of continuity with the traditions in philosophy of mind and philosophy of science, and it is in no way essential for the position defended in this thesis: if pluralism is the right approach, it is right regardless of whether or not there is a clear distinction between psychology and neuroscience.

[3] A distinction is sometimes made between neurophilosophers and philosophers of neuroscience. Neurophilosophers (e.g., Patricia Churchland, John Bickle) apply findings from neuroscience to traditional philosophical problems, such as free will or consciousness. Philosophers of neuroscience (e.g., William Bechtel, Carl Craver) consider traditional problems of philosophy of science with regard to neuroscience. My approach differs from both of these and is somewhere in between. I apply results and insights from philosophy of neuroscience (and philosophy of science in general) to address traditional problems in philosophy of mind.

than that of other philosophers of the same tradition (for example, Ned Block, David Chalmers, Frank Jackson, or Joseph Levine). Although I am criticizing Kim, it is beyond doubt that his contributions to philosophy of mind have been groundbreaking. I greatly admire him for the clarity and beauty of his philosophy and few philosophers have influenced my intellectual development as much as he has.

While I was already halfway through writing this book, I came across an excellent recent work with aims strikingly similar to mine: Steven Horst's (2007) *Beyond Reduction*. Horst is also arguing against reductionism, defending pluralism, and emphasizing the importance of bringing philosophy of mind closer to philosophy of science. Fortunately, there are also substantial differences in our arguments and conclusions. In contrast to Horst, mechanistic explanation and the interventionist account of causation play a key role in my arguments, and the "cognitive pluralism" of Horst is more far-reaching and radical than the pluralistic physicalism I am defending. Furthermore, Horst does not discuss the functional model of reduction, which receives a lot of attention in this thesis. On the other hand, he goes far deeper into the details of some other debates in the philosophy of mind, most importantly the debates on supervenience and the "explanatory gap." Therefore, although the spirit of Horst's book is very close to that of this one, the two are considerably different and complementary contributions to philosophy of mind.

The structure of this thesis is as follows. In Part I, I will discuss reduction and reductionism in philosophy of science, focusing on psychology and neuroscience. I will go through the problems of the classic intertheoretic models of reduction and the more recent "ruthless" approach to reductionism, and defend a position consisting of two main elements: *mechanistic explanation* and the *interventionist account of causation*. This leads to *explanatory pluralism* regarding psychology and neuroscience. In the end of the part, I will also consider the issue of levels and its relation to reduction.

In Part II, I will criticize the way reduction has been understood in philosophy of mind, based on what has been presented in Part I. I will go through classical topics like multiple realizability, functionalism, the explanatory gap, and nonreductive physicalism, and show how our

understanding of them is changed once we have a proper picture of reduction. An extensive and detailed section is devoted to criticizing the functional model of reduction, which has become something like a standard model in philosophy of mind.

In Part III, I will present and defend a new framework for philosophy of mind. Its main elements are explanatory pluralism, mechanistic explanation, and the interventionist account of causation. I will also develop an ontological framework for this position, which consists of a kind of ontological pluralism based on the idea of robustness. Subsequently, I will show that the causal exclusion argument does not make this position incoherent, and that the position is compatible with certain forms of physicalism, to the extent that it could be called *pluralistic physicalism*. In the end, I will argue that many reductionist ideas fit perfectly into this pluralistic framework, including for example the thesis that all mental properties can be mechanistically explained.

PART I:
Reduction in Philosophy of Science

Introduction

In this part, I will discuss reduction[4] as it has been understood in the philosophy of science of the 20th (and 21st) century. Going through the history (or prehistory) of reductionist ideas would be interesting, but this thesis is not a historical one, and therefore I will only discuss the models that are most relevant to contemporary debates. I will begin with the development of intertheoretic models of reduction that started in the 1950s, in the afterglow of logical positivism, and then go on to discuss more recent accounts of reduction, most importantly "New Wave Reductionism" and "Ruthless Reductionism." I will argue that these approaches face fatal problems, at least in the case of psychology and neuroscience, and that "mechanistic explanation," especially when supplemented with the interventionist account of causation, provides a more accurate and scientifically credible framework for approaching issues of reduction. In the end, I will consider the question of levels and its relation to reduction, focusing on the problems in current accounts of levels.

[4] Throughout this thesis, I use the term "reduction" to refer to a single case of accomplished or purported reduction: the reduction of thermodynamics to statistical mechanics, the reduction of chemistry to physics, and so on. "Reductionism" refers to a broader thesis, according to which reductions are to be expected (a predictive claim) and/or desirable (a normative claim). Of course, different models of reduction yield different reductionisms, and one can be reductionist regarding some domains of science but not others. Therefore, for instance, "psychoneural Nagel reductionism" means the thesis that psychology will be or should be reduced to neuroscience following Nagel's model of reduction.

1. Reduction: From Derivations of Theories to Ruthless Metascience

By far the most influential philosophical models of reduction have been the "intertheoretic" models, where reduction is seen as a relation between formal theories. The development of intertheoretic models started in the middle of the 20th century, drawing on the spirit of logical positivism. The ultimate goal was to show how unity of science could be attained through reductions. John Kemeny and Paul Oppenheim (Kemeny & Oppenheim 1956) formulated reduction as a relation between theories, where the reducing theory should be able to explain any observational data that the reduced theory explains, and the reducing theory should be at least as well systematized[5] as the reduced theory. A few years later, Oppenheim and Putnam published their extremely influential "Unity of Science as a Working Hypothesis" (1958), where they presented the hypothesis that all sciences will be reduced to the fundamental physical science via "microreductions." In a microreduction, the higher-level entities to be reduced must be fully decomposable into the reducing entities of lower levels. Oppenheim and Putnam also adopted the conditions for reduction stated by Kemeny and Oppenheim (1956). That is, according to Oppenheim and Putnam, a theory T_2 *microreduces* to theory T_1 if and only if (1) any observational data explainable by T_2 are explainable by T_1, (2) T_1 is at least as well systematized as T_2, and (3) all the entities referred to in T_2 are wholes which are fully decomposable into entities in the universe of discourse of T_1. This is in effect a model of *replacement*, since the successful microreduction makes T_2 entirely dispensable. This account suffers from serious defects that I will only briefly mention here (see, e.g., Sklar 1967 for more details): it assumes that we can clearly distinguish between observational and non-observational terms, the notion of systematization or systematic power is not clearly defined, and it is hard to find examples from history of science that would satisfy the requirements.

[5] A theory is well systematized if it is simple but predicts or explains a broad range of phenomena. That is, systematization or systemic power is a measure that combines simplicity and strength. Kemeny and Oppenheim acknowledge the need for a more precise definition, but do not give one in the paper.

Also in Nagel's (1951; 1961, 336-397) classic account of reduction, many ideas of logical positivism are clearly visible. Reduction is seen as a relation between formal theories, such that the theory to be reduced (T_2) is logically derived from a more fundamental theory (T_1). Conditions for a successful reduction are that (1) we can connect the terms of T_2 with the terms T_1, and that (2) with the help of these connecting assumptions we can derive all the laws of T_2 from T_1. In Nagel's model, a reduction can be seen as a kind of deductive-nomological explanation, where T_1 explains T_2.

Nagel distinguished between two different kinds of reductions: "homogeneous" and "heterogeneous" reductions. In a homogeneous reduction the two theories share the same conceptual apparatus. For example, the reduction of Galileo's laws to Newtonian mechanics was a homogeneous reduction. However, most (interesting) cases of reduction are heterogeneous reductions, where one of the theories has concepts not found in the other. In these cases, in order to satisfy the two conditions for reduction, we need some principles or laws that connect the terms of the two theories. The exact nature of the connecting principles, or "bridge principles/laws" as they came to be called, was left open by Nagel, and has been a matter of much debate. Although the conditions of a Nagel-type reduction can be fulfilled already when these laws express material conditionals of the form "$\forall x\ (F_{T1}x \rightarrow F_{T2}x)$" (e.g., Richardson 1979), it was widely accepted that biconditionals of the form "$\forall x\ (F_{T1}x \equiv F_{T2}x)$" are necessary for the ontological simplifications that were considered to be one of the main goals of reduction.

Nagel presented the reduction of thermodynamics to statistical mechanics as a paradigmatic example of a successful scientific reduction. He focused on the derivation of the Boyle-Charles' law for ideal gases ($pV = kT$, where p is the pressure of the gas, V is the volume of the gas, T is the absolute temperature of the gas, and k is a constant) from statistical mechanics, pointing out that the derivation of the whole thermodynamics would be immensely complicated, and that even for the derivation of the Boyle-Charles' law many idealizing assumptions have to be made: one has to assume, for example, that the gas is composed of a large number of perfectly elastic spherical molecules with equal masses and volumes but with dimensions that are negligible compared to the distances between the